The C. elegans Lifespan Machine and its application to the temperature scaling of lifespan

The *C. elegans* Lifespan Machine and its application to the temperature scaling of lifespan

A dissertation presented

by

Nicholas Edward Stroustrup

to

The Committee on Higher Degrees in Systems Biology

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Systems Biology

Harvard University
Cambridge, Massachusetts

April 2013

Professor Walter Fontana Nicholas Edward Stroustrup

The *C. elegans* Lifespan Machine and its application to the temperature scaling of lifespan

Abstract

Lifespan results from the complex interaction between genetic, environmental and stochastic factors, and therefore varies widely even among isogenic individuals. In *C. elegans*, the action of molecular mechanisms on aging can be inferred from their statistical effects on the distribution of lifespans within populations. However, such investigations are hindered by limitations in the methods available for collecting lifespan data. To enable the rapid collection of survival curves at any desired statistical resolution, we developed an automated platform for determining the lifespans of large populations of nematodes.

Our method combines high-throughput imaging with automated image analysis to generate a visual record of individual nematode deaths, from which survival curves are automatically constructed and visually validated. This approach produces results consistent with the manual method for several mutants in both standard and stressful environments, allowing rapid screening for genetic and environmental determinants of lifespan and enabling quantitative investigations into the statistical structure of aging.

We applied our method to study the effects of temperature on *C. elegans* aging across the range of 20 °C to 36.5 °C at fractional degree intervals. We found that the functional form of lifespan distributions appears invariant to temperature, and is well-described across the entire temperature range by a Weibull model including a frailty correction. *C. elegans* mean lifespan, however, appears not to follow a single scaling function of temperature. Rather, lifespan scales with temperature in a segmental manner, suggesting distinct physiological transitions occur around 30 °C, 31.5 °C, and 35 °C. These findings imply a distinction between the molecular mechanisms that determine average lifespan, which differ between temperature ranges, and the overall stochastic behavior of mechanisms generating variation in lifespan, which appear temperature invariant.

We integrate these findings to propose a model of *C. elegans* aging where death occurs as a consequence of the first failure among vital components, all failing independently according to the same temporal distribution. Extreme value phenomena such as this can produce Weibull distributions across a diverse range of failure dynamics in underlying components, and thus might explain the shape of lifespan distributions and its conservation across a wide temperature range.

To my parents who raised me in the woods, to Debbie who showed what success looks like and what it requires, to Javier who taught me his love and respect for the worm, to Walter who provided mentorship, support, and shared his uncompromising aesthetic for research, to Remy, Martin, Kalin, Yifat, Pedro, Daniel, Tom, Dan, Cata, and JB who brought style to the madhouse, to Annemarie and Kris for their boldness, and to Norah, Gavin, Courtney, and Anya for their boundless and infectious enthusiasm.

This project would have been impossible without the combined efforts of Bryne Ulmschneider (BU), Zachary Nash (ZN), Isaac F. López Moyado (ILM), Cam Watson (CW), and Winston Anthony (WA) who during their time in the Fontana Lab executed many of the experiments described in this document. Their individual efforts are detailed in the captions of each figure, but this represents only a small fraction of their overall contribution.

The conservatism of early ontogeny, expressed as the various modes of cell division, cell differentiation, and morphogenic movement, become penetrable to standard neo-Darwinian thinking with one simple modification: evolutionary pattern has arisen not by selection on individuals alone, but by the interactive effects of selection operating at differing levels of biological organization

Leo W. Buss

Table of Contents

Chapter 1

Aging and the Measurement of Lifespan

Aging organisms exhibit functional declines at many levels of biological organization, spanning molecules, cells, tissues, and organs. These declines ultimately conspire to kill the organism. A close look at the tissues of an eighteen-day-old *C. elegans* nematode leaves one unsurprised it is near death. Electron micrographs reveal muscle sarcomeres in disarray, a distorted and discolored pharynx, and a cuticle both ridged and thickened [Herndon et al., 2002]. This structural disorder is mirrored by molecular dysfunction-transcriptional regulation is perturbed [Budovskaya et al., 2008], protein homeostasis disrupted [David et al., 2010], subcellular organelles compromised [McGee et al., 2011], and somatic copies of the genome are perfused with mutations [Dolle et al., 2000, Maegawa et al., 2010]. This catalogue of ailments is not specific to nematodes-—most metazoans demonstrate similarly complex, multi-faceted structural and molecular declines, including mice [Boaro et al., 1998], zebrafish [Wuhr et al., 2011], humans [Fuster et al., 1992], and even marine invertebrates [Loram and Bodnar, 2012].

In a sense, the startling observation is not that complex biological systems fail over time, but that they exist at all. The molecules, cells, tissues and organs that fail with age exist only as a startling outcome of evolutionary processes. Complex biological structures are created and maintained across generations because variants are selected for their ability to increase the fitness of the

individual in which they function. More than half a century ago, Haldane, Medawar, and Williams observed that such selection might be age-dependent, insofar that phenotypes presenting late in the life of an organism may contribute less to its fitness than phenotypes presenting early [Rose et al., 2008]. Deleterious mutations that, for example, might cause a pharynx to collapse or an epidermis to wrinkle, could be expected to accumulate without consequence if such symptoms manifested after an individual loses relevance to its children, as is often the case after reproduction ceases. Regardless of reproduction, extrinsic mortality will be an increasingly dominant determinant of fitness–the cumulative probability of an accident, environmental fluctuation, or predation killing an organism will increase over time. In all but a perfectly safe environments, the likelihood of a accidental death will ultimately reach certainty, setting a time-scale on which even the most robust and resourceful organism will have no hope of surviving [Austad et al., 2007]. The failure of complex systems, in a sense, reveals the limits of evolution's ability to create and maintain structured complexity.

Biological systems should therefore be *expected* to unravel, as continuing functionality becomes increasingly unlikely to contribute to fitness and selective pressures therefore increasingly fail to maintain order. This unravelling need not have any universal structure or even any characteristic mechanistic basis. Aging might not even be a biological process in the traditional sense, instead representing the emergent consequence of diminishing relevance of evolutionary processes over the lifespan of an individual.

Though aging itself need not be under direct evolutionary selection, it clearly involves the decline and failure of evolutionary-conserved components. A quarter of US residents die because their heart fails [Heron, 2012], and their hearts as evolved components will show failure dynamics determined by the nature of the heart's evolutionarily-conserved structures. A bridge fails according to the nature of its steel and concrete construction; a heart will fail as a result of its molecular and cellular construction. Clearly, a better understanding of such failures is of importance to the researcher who would like to understand the basic principles involved, and to the clinician who would like to slow aging down or alleviate its consequences. Yet, the multi-factorial nature of aging makes it inherently difficult to distinguish aging's causal factors from its symptomatic consequences. How can a molecular process observed to change over time

be assigned a causative rather than incidental role? In a system where evolution may not be acting to produce coherent, explainable phenomena, how can a researcher separate the important molecular mechanisms from a host of irrelevant pathological detail? A pervasive strategy for doing so involves the measurement lifespan. In controlled environments, death is an easily observed, relatively unambiguous consequence of aging processes. In so far as aging processes set the timing of death, molecular perturbations–mutations, environmental changes, exposure to compounds–can be evaluated for their effect on aging by their effect on lifespan. Measurement of the timing of death provides a quantitative phenotype, a relatively unambiguous measure by which molecular perturbations can be evaluated for their role in aging.

1.1 The molecular determinants of aging are evaluated relative to their effect on lifespan

Work by Michael Klass [Klass, 1977]), Tom Johnson [Friedman and Johnson, 1986], and Cynthia Kenyon [Kenyon et al., 1993] provided early evidence suggesting *C. elegans* lifespan might be determined by a small number of genetic loci. The cloning of *daf-2* [Kimura et al., 1997] and *daf-16* [Lin et al., 1997], and the subsequent demonstration of the insulin/IGF signaling pathway as an evolutionary conserved determinant of lifespan among metazoans [Tatar1 et al., 2001, Taguchi et al., 2007], redefined the study of aging as one accessible by modern molecular biologic approaches. The very existence of single genes acting to shorten lifespan presented a challenge not only to molecular geneticists, who aimed to understand the mechanisms downstream of the signalling pathways discovered, but also to evolutionary biologists now needed to explain how such mechanisms might exist despite their apparent conflict with the somewhat nihilistic existing evolutionary theory.

Several approaches have been taken to follow up on the discovery of conserved, lifespan-extending signalling pathways, with most ultimately based on molecular inquiry followed by an evaluation of uncovered mechanisms effect on lifespan. Regardless of whether a molecule was discovered through a classic genetic screen or modern proteomic approaches, the first or last figure of the paper depicts a survival curve of a perturbed population. In this way,

characterizations of the insulin/IGF signalling cascade, originally mapped entirely based on its regulation of *elegans*' entry and exit from diapause, [Larsen et al., 1995], could be translated directly into a list of candidate determinants of lifespan. Unbiased RNA-interference screens [Lee et al., 2002, Hamilton et al., 2005, Hansen et al., 2005, SP and G., 2007] or screens designed specifically to identify insulin signalling components [Samuelson et al., 2007], provided a bounty of lifespan determinants, as did independent efforts to identify transcriptional and translational targets of daf-2 and daf-16 [Murphy, 2005, McColl et al., 2010, Oh et al., 2005].

1.2 The growing importance of quantitative measurements of lifespan

In the author's opinion, the growing list of *C. elegans* genes known to effect lifespan paints a picture where a large set of biological components act quantitatively in concert to determine lifespan. Though many significant factors downstream of *daf-16* have been identified, no single factor accounts for more than a fraction of the lifespan extension granted by *daf-2*. It is clearly the case that alterations to *elegans* signaling networks produce sensational lifespan extensions of 100% or more. However, it is possible that disruption of actual aging mechanisms, were they truly as numerous and diffuse as current data suggests, will never produce such a clear result. In this case, the study of aging mechanisms might necessarily require significantly more quantitative reasoning than current experimental procedures permit. The effect of a gene on lifespan might not be well summarized by whether or not it extends lifespan, but by exactly how much and to what extent its action couples quantitatively with the action of other genes. If true, this represents a major technical problem for the field. For all its sensitivity, lifespan assays are notoriously irreproducible [Lee, 2006], time-consuming, and monotonous. These issues are relatively minor when studying dramatic 50% or 100% lifespan extensions, but for the characterization of more subtle determinants, technical issues become problematic and, in many cases, prohibitive.

Many challenges exist in the collection of lifespan data, but one limitation is paramount. The nature of manual lifespan assays make it technically and emotionally challenging to run sufficient replicates to accurately interpret small effects. In chapter 3, we demonstrate lifespan varies by

almost 18 percent in subsequent replicates of the same wild-type strain in controlled conditions using standardized, automated equipment. Such uncontrolled environmental variation can be partially mitigated through careful experimental design–usually involving the measurement of perturbations relative to an internal control. In this way, lifespan extensions are quantified relative to wild-type animals are run simultaneously. However, in most experiments it is necessary to house different strains on separate plates (as strains often look identical and cannot be labelled or distinguished quickly by eye). This opens the possibility for environmental differences observed between experiments to exist within different strains handled in a single experiment. Furthermore, there is no guarantee that perturbations under study do interact epistastically with the uncontrolled variation. The lifespan extension of, say, *daf-2(e1368)* might vary significantly relative to wild type in subsequent replicates. These problems conspire to reduce the quantitative quality of lifespan experiments, and in the presence of small effects can generate lots of spurious results. The obvious solution for this is to run many replicates of lifespan experiments, allow the experimenter to characterize inter-replicate variation, and demonstrate the invariance of a a perturbation across replicates. However, because lifespan assays are time-consuming and monotonous, such replicates are rarely performed in practice. Such problems will only become worse as increasingly smaller effects are studied, and curtailing the reproducibility and scope of data collected.

1.3 Automation as a means for increasing the quality and scope of *C. elegans* demographic data

To address the growing need for better lifespan data in *C. elegans*, we have developed a platform for automated acquisition of nematode survival curves. The method is designed to be scalable and relatively inexpensive, providing a standardized method for producing survival curves at arbitrary statistical resolution. It becomes possible to collect lifelong observations of nematode populations ranging from thirty to more than three thousand animals with similar accuracy and precision. Our method combines groups of flatbed scanners to function as a single spatially extended microscope. Alone or in groups, these scanners produce time-lapse videos under highly controlled environmental conditions. We couple this imaging approach with an image-analysis

pipeline that automatically identifies death times. By creating an auditable trail of images, the system enables the rapid validation and refinement of processed data, and thus the statistically rigorous analysis of aging and stress-resistance phenotypes. The technology is described in detail in chapter 2 and in the accompanying technical supplement. In chapters 3 this method is used to investigate the statistical form of *C. elegans* lifespan distributions, and in 4 the temperature dependence of this statistical form is used to provide a hypothesis as to the underlying stochastic processes at work.

Chapter 2

The *C. elegans* Lifespan Machine

The text and figures for this chapter are taken from a manuscript of the same name. A version of the manuscript lacking Fig. 2.5 is currently in press at the journal Nature Methods. The authors of this manuscript are Nicholas Stroustrup, Bryne E. Ulmschneider, Zachary M. Nash, Isaac F. López Moyado, Javier Apfeld, and Walter Fontana. The author contributions are as follows—N.S. designed and implemented hardware and software. N.S. and B.U. constructed and calibrated equipment. N.S. and J.A. conceived and designed experiments. N.S., B.U., J.A., Z.N. and I.F.L.M. performed experiments. N.S. designed analytic tools. N.S., J.A. and W.F. provided guidance, analyzed data, interpreted results, and wrote the manuscript. J.A and W.F. are co-last authors.

2.1 Introduction

Aging organisms exhibit functional declines at many levels of biological organization. These declines ultimately conspire to kill the organism and determine its lifespan. Because lifespans vary remarkably [Kirkwood et al., 2005] even within isogenic populations, useful information about aging mechanisms can be obtained by identifying their statistical signature in the response of lifespan distributions to genetic, chemical and physical interventions

C. elegans individuals live as self-fertile adults for a few weeks, producing large numbers of isogenic offspring. Early investigations revealed single point mutations in insulin/IGF-1 pathway components capable of doubling average lifespan [Friedman and Johnson, 1986, Kenyon et al.,

1993]. These findings subsequently generalized to other organisms, including fruit flies and mice [Kimura et al., 1997, Kenyon, 2010], establishing *C. elegans* as a metazoan model for studying the genetics of aging. The acquisition of survival curves in *C. elegans* thus became an essential part of aging research. In routine practice, survival curves of animals cultured on solid agar in a Petri dish, fed by a bacterial lawn of *E. coli*, are acquired manually through daily observation with a low-power dissecting microscope [Sutphin and Kaeberlein, 2009, Wilkinson et al., 2012]. Death is recognized by the failure of an individual to react by motion to prodding with a wire, requiring labor-intensive, repetitive, and subjective observation. This protocol provides strong incentives to keep population sizes small, observations infrequent, and the number of replicates limited, thus curtailing the reproducibility and scope of data collected.

We present a scalable method for producing standardized survival curves at arbitrary statistical resolution, allowing "hands-free", lifelong, and accurate observation of arbitrarily large nematode populations. Our method combines groups of flatbed scanners to function as a single spatially-extended microscope. Alone or in groups, these scanners produce time-lapse videos under highly controlled environmental conditions. We couple this imaging approach with an image-analysis pipeline that automatically identifies death times. By creating an auditable trail of images, the system allows rapid validation and refinement of processed data, and therefore supports a statistically rigorous analysis of aging and stress-resistance.

In several cases, notably in worms and flies, focused efforts have been made at acquiring high-resolution survival data [Curtsinger et al., 1992, Carey et al., 1995, Vaupel et al., 1998, Vanfleteren et al., 1998, Johnson et al., 2001, Mair et al., 2003, Baeriswyl et al., 2009, Wu et al., 2009]. Our method formalizes, standardizes, and routinizes such observations for nematodes, enabling the accumulation of a self-consistent body of quantitative demographic data tightly linking aging, genotype, and environment. No such comprehensive body of data exists for any metazoan to date, humans included. We refer to the integrated combination of scanner hardware and image-processing software as the "Lifespan Machine" (LM).

2.2 Results

2.2.1 A distributed microscope for large worm populations

Lifespan data are acquired over many scales, ranging from aging studies spanning multiple months to stress-resistance assays lasting less than a day [Lithgow et al., 1995], involving targeted characterizations of only a few strains [Johnson et al., 2001] or screening projects of thousands of gene knock-downs [Friedman and Johnson, 1986, Kenyon et al., 1993, Kenyon, 2010]. Throughput is important for screening projects, while targeted studies benefit from large populations to provide statistical power. To suit all these purposes, our platform combines flatbed scanners in a scalable fashion to monitor arbitrarily large populations over a wide-range of timescales.

Our approach replaces human observation with an automated system that acquires and processes time-lapse images to identify the timing of death defined as the persistent cessation of spontaneous movement. Our current reference implementation can simultaneously monitor 30,000 animals across 800 plates, imaging 4.5 m^2 of agar surface at a resolution of 8 μm (Supplementary Notes 1—4). This capacity is used to run many independent, single-scanner experiments in parallel with larger investigations involving ten scanners or more.

Young worms move quickly relative to the observation frequency attainable with scanners, precluding their tracking. However, as animals slow down with age, their motion can be quantified longitudinally, especially when it is limited to changes in posture at a fixed location.

Each scanner monitors 16 Petri dishes (plates) sealed face down to a glass sheet by a rubber mat, with each plate containing a population of about 35 animals (Online Methods). A fluorescent lamp inside the scanner lid sends light through plates, agar, bacteria, and worms, to be captured by a sensor chip moving underneath the glass surface in synchrony with the light source (Fig. 2.1a, Supplementary Notes 1 and 3). A scanner thus acts as 16 camera-equipped dissecting microscopes, while costing significantly less and in a form factor suited for dense shelving inside temperature-controlled incubators.

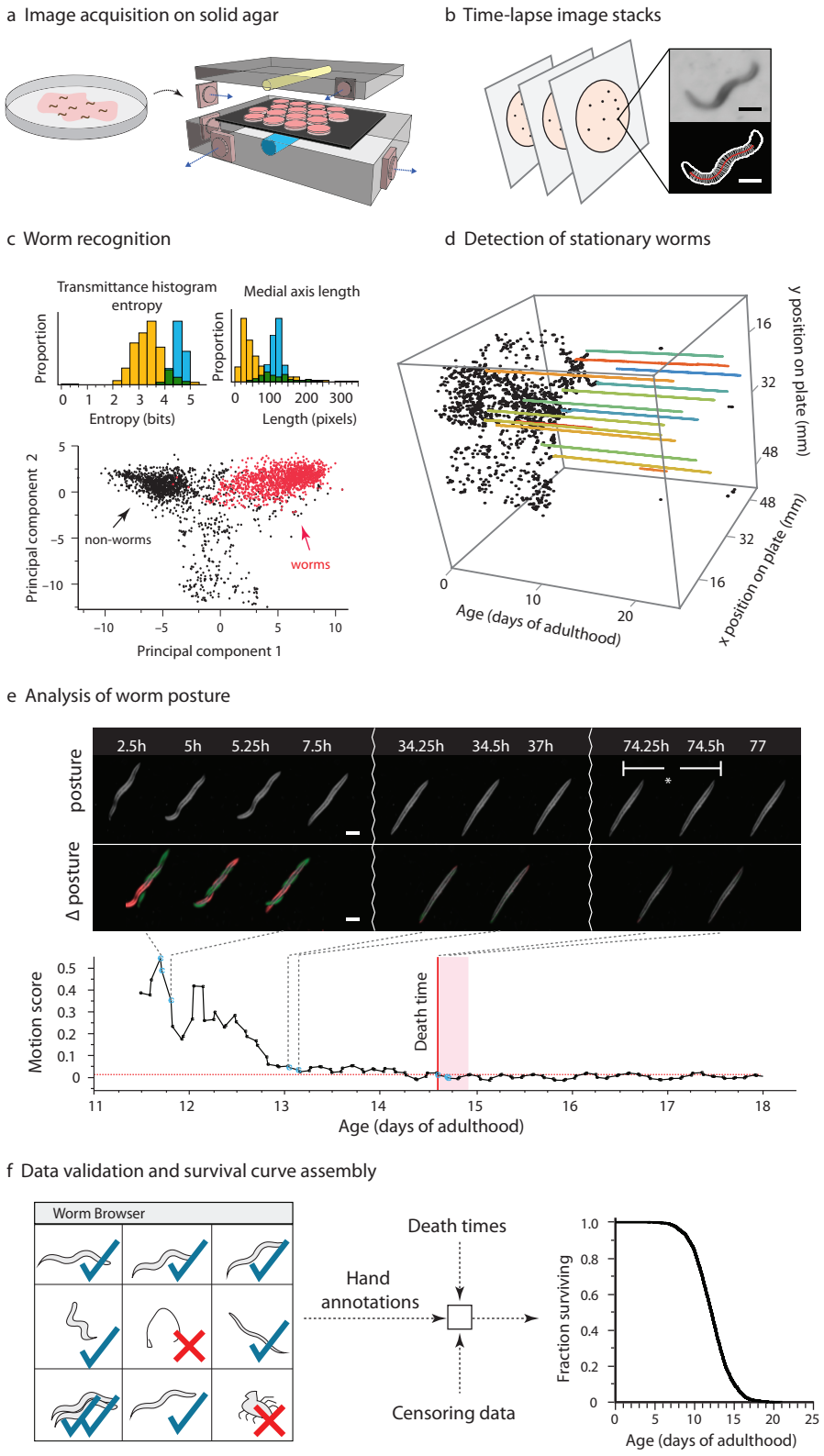

Figure 2.1: Lifespan Machine workflow. Caption continued on next page.

Figure 2.1 (Cont'd): Lifespan Machine workflow.(a) Agar plates with age-synchronized worms are placed face-down on the surface of flatbed scanners (Supplementary Note 1). **(b)** Each scanner captures a time series of images of each of its plates. Images are processed to identify foreground objects (Supplementary Note 2). Scale bar = 250 μm. **(c)** Various morphological features are quantified for each object. The population distributions of these features demonstrate that none permits by itself a discrimination of worms from non-worms. Based on many validated images of worm and non-worm objects, classifiers are constructed that are capable of categorizing objects based on their 65-dimensional feature vectors. The panel depicts a projection of the point cloud representing the validated set of feature vectors, indicating that differentiating worms (red) from non-worms (black) is possible (Supplementary Note 2). **(d)** Once worm objects have been identified, individual image stacks are analyzed to identify animals that have become stationary (Supplementary Note 5). **(e)** Stationary animals are further analyzed to detect posture changes, including head and tail movements. Death is identified by retrospective analysis as the final cessation of postural change (Supplementary Note 5). **(f)** The LM produces a time-lapse image record for each individual. These records can be used for visual validation or resolution of ambiguities employing a rapid data-inspection tool, the "Worm Browser" (Supplementary Video 6 and Note 4). Validated death times are combined with automatic censoring data (Supplementary Note 8) into a Kaplan-Meier survival curve (actual data shown) or into hazard rates using a statistics package.

Traditional protocols for monitoring large populations of nematodes over long periods of time move plates in and out of an incubator [Sutphin and Kaeberlein, 2009, Wilkinson et al., 2012, Vanfleteren et al., 1998, Johnson et al., 2001], exposing them to spatially and temporally heterogeneous conditions. Our apparatus maintains plates at a fixed position within a controlled environment for the duration of each experiment. This is important for accuracy, given the exquisite temperature sensitivity of *C. elegans* lifespan [Klass, 1977] (Supplementary Note 1). We curbed temperature excursions by two systems of fans, one cooling each scanner, the other circulating air evenly throughout the enclosing incubator (Supplementary Note 3). Temperature fluctuations on a typical scanner surface were thus reduced from several degrees to less than 0.5 °C, while limiting differences between scanners to about 1 °C (Supplementary Note 1).

To improve image quality we repositioned a single fixed lens in the scanner's optical path, shifting the focal plane a few millimeters above the scanner glass, matching the agar surface inside plates. To ensure that the agar surface is consistent across experiments, we developed a robust method for the controlled drying of agar plates (Online Methods).

2.2.2 Automated identification of death times

Following an optimized scanning schedule (Online Methods), the LM captures 24 images of each plate per day. The image-processing pipeline identifies worm objects (Supplementary Note 2) and their position in every frame (Fig. 1b,c). Individual animals, once stationary (Fig. 1d), continue to change posture for a time, often moving their head or tail (Fig. 1e, Supplementary Videos 1-5 and Note 5), reminiscent of the "class C" animals described in studies of age-dependent locomotory decline [Herndon et al., 2002]. Our software determines the final cessation of spontaneous postural movement through retrospective image analysis (Online Methods, Supplementary Note 5 and Video 6). Posture analysis of stationary animals is crucial to avoid underestimation of lifespan. We observed wild type animals spending on average 1.2 days of their life in this state (Fig. 1e), while certain mutants, for example age-1(hx546), persist in it more than twice as long (Supplementary Note 6), consistent with previous reports [Huang et al., 2004].

We found that most animals exhibit a stereotyped morphological change at or near the time of their final posture change (Supplementary Note 7). Individuals first shrink by at least 10% and expand by more than that amount shortly thereafter. We never observed a change in posture after this expansion. The correlation between the cessation of postural motion and this morphological change suggests a physiological transition, perhaps corresponding to the animal's death, corroborating our death criterion.

A crucial component of our platform is a software package that allows rapid validation of death times through visual inspection (Supplementary Note 4). This quality control step (Fig. 1f, Supplementary Video 6, Online Methods) allows a user to determine whether plates were compromised (e.g. desiccation or fungal contamination) and need exclusion. In particular, it permits identification of aggregates containing multiple worms (Supplementary Note 8) and the censoring of incorrectly classified objects (e.g. features of the agar lawn) and non-aging related deaths (e.g. body rupture), thus enabling a visual validation of machine operation.

2.2.3 The LM produces accurate and precise survival curves

To evaluate whether the LM produces results consistent with those of a human observer, we placed roughly 8,000 age-synchronous wild type animals into a single liquid suspension, and distributed aliquots of approximately 35 individuals across 170 agar plates seeded with *E. coli* OP50 (Online Methods). 160 of these plates were distributed across ten scanners situated in an incubator at 25 °C. The remaining 10 plates were placed in a separate incubator and monitored daily by hand using a dissecting microscope.

Because we drew all animals from the same pool, any significant difference between survival curves of individual plates must be attributed to either measurement error or environmental variation, such as differences in plate temperature. We compared death times grouped by plate on a single scanner with those acquired by the manual procedure (Fig. 2a). In both methods, plate means varied within essentially the same range, between 10.7 and 13.1 days. No significant effect was observed between plates collected by the manual method (Log-rank test for homogeneity P =0.059). A small, significant difference was observed between plates collected by the automated method (Log-rank P =0.002), reflecting either inaccuracies in the estimation of death times or a slightly increased environmental variability among plates. In two independent replicates, the survival curves determined by the LM differed in mean lifespan by 2.5 hours or 0.2% (Log-rank P =0.01, Fig. 2b) and 11.7 hours or 3.2% (Log-rank P <0.001) from those of cohorts assayed using the conventional method. These small differences could be explained by practical limitations in our ability to measure and maintain temperature constant in both methodologies.

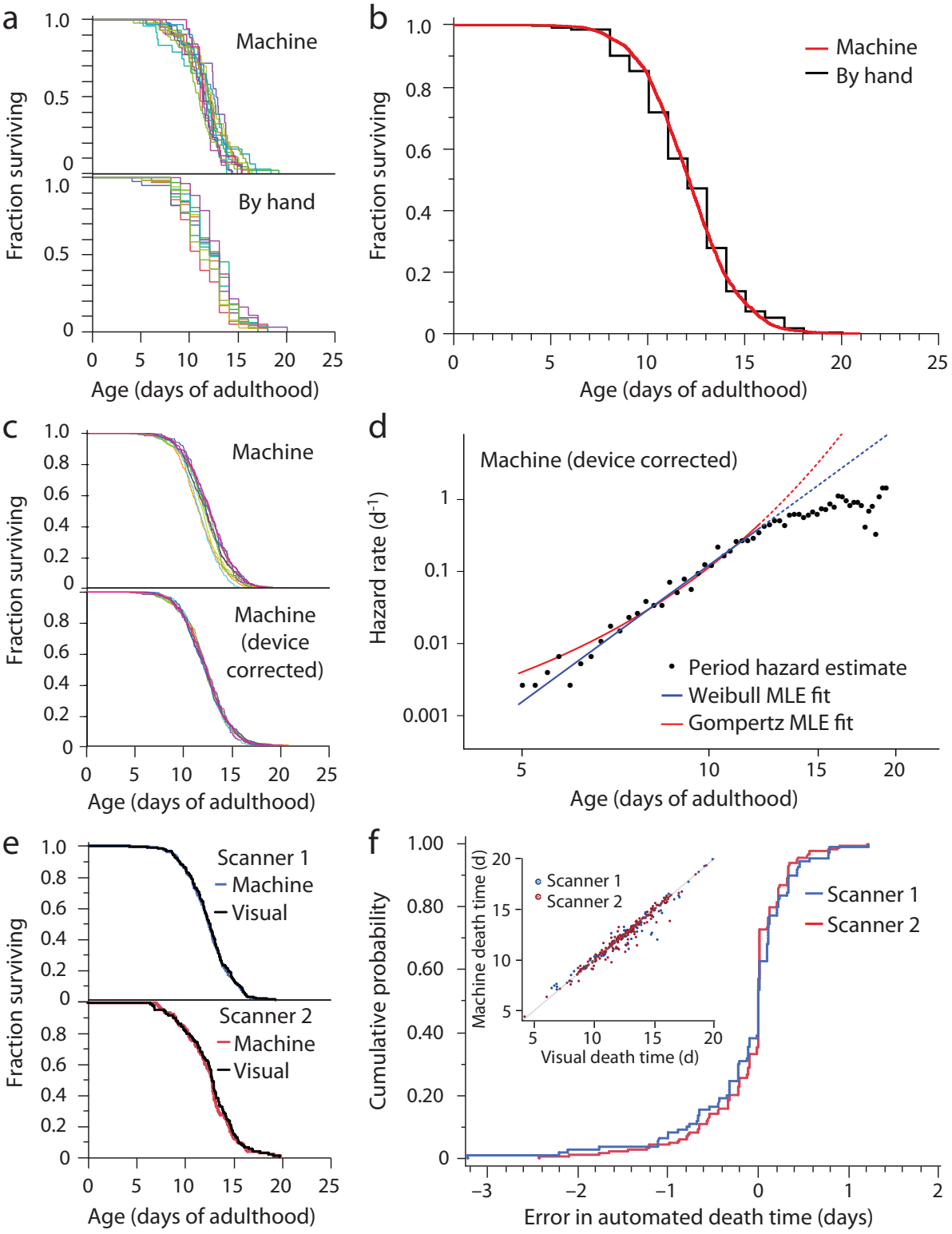

Figure 2.2: Automated wild type survival data. Caption is continued on the following page

Figure 2.2 (Cont'd): Automated wild type survival data.(a) A survival curve was generated from 484 death times of wild type animals co-located on a single scanner (top) and 513 wild type animals observed using the manual method (bottom). **(b)** The entire population of 3,578 wild type animals over ten scanners was aggregated into a single curve, nearly indistinguishable from the manually-scored curve. **(c)** Scanner-specific microenvironments, mainly differences in temperature, affect lifespan (top). The effect was estimated using a categorical Accelerated Failure Time model, to obtain device-corrected lifespans registered to the grand mean (bottom), as detailed in Supplementary Note 10. **(d)** Hazard rates were estimated from lifespan data (Supplementary Note 11). The blue (Weibull) and red (Gompertz) lines represents a maximum likelihood fit of the parametric model using data up to median survival. A smaller automated assay performed at 20 °C on HT115 bacteria was consistent with the mortality kinetics at 25 °C on OP50 (Supplementary Note 11). **(e)** Image records of populations on two scanners were selected for validation using the Worm Browser to determine death times by eye (Supplementary Video 6 and Note 4). The resulting curves compare with those produced by the automated method (Log-rank P > 0.5 in both cases). **(f)** The cumulative LM error (i.e. the death time differences between automation and visual inspection) is shown for both scanners. Visual and automatic lifespans are highly correlated (R^2 = 0.96), as shown in the inset, with the LM underestimating lifespan by an average of two hours. Experiments performed by BU and NES.

Where experiments require comparison between populations much larger than 300 individuals, animals must be distributed across multiple scanners. With the statistical power (Supplementary Note 9) afforded by each scanner, even slight environmental differences between scanners translate into statistically significant differences between survival curves (Fig. 2c top). Indeed, scanner surface temperature correlated well with mean lifespan (R2 = 0.75; P=0.023; Supplementary Note 1).

Because devices produced a small effect on lifespan, we applied an accelerated failure time regression model with scanner identity as a categorical covariate to generate "device-corrected" lifespans (Fig. 2c, Supplementary Note 10). This procedure is useful for assessing whether data pooling is justified and for standardizing data against a desired baseline curve, for example, aligning large experiments to a single scanner.

The LM provides data in quantity and quality appropriate for estimating time-dependent hazard (mortality) rates, Supplementary Note 11. We observed a rapid increase and subsequent deceleration in mortality with time, as noted previously 10-12,14. For the first 50% of deaths, mortality appears better fit by a power of time t (Weibull hazard, $\frac{\alpha}{\beta}(\frac{t}{\beta})^{(\alpha-1)}$ than an exponential

in time (Gompertz hazard, a et/b), Fig. 2d and Supplementary Note 11, though for values of the Gompertz parameter a much smaller than 1, Gompertz and Weibull distributions behave very similarly for all but the lowest quantiles of survival, making disambiguation difficult. An analysis of the full hazard data, accounting for the deceleration phase, may require frailty models to incorporate phenotypic diversity potentially present even in isogenic populations [Vaupel et al., 1998, Weitz and Fraser, 2001a]. However, the observed hazard deceleration is not the result of environmental heterogeneity in our apparatus, as it persists after device-correction (data not shown).

The close agreement between automated and manual survival curves was surprising, as scanners expose animals to conditions that differ from the manual method, such as oscillating lights [Mathew et al., 2012] and temperature oscillations (Supplementary Note 1). The difference in lifespan between animals exposed and not exposed to the scanner environment, but scored by hand, was so small as to be explained by limitations in our ability to maintain the same temperature in both environments (Supplementary Note 12). Importantly, the LM and the manual procedure adopt different operational definitions of death—cessation of spontaneous vs. stimulated movement, respectively. We therefore sought a validation process other than comparison to the manual method by developing software to assist in the visual inspection of the image record generated by the LM for each experiment (Supplementary Note 4 and Video 6). Two survival curves constructed from 294 user-annotated death times based on images from two scanners were each statistically indistinguishable from curves produced by automated annotation (Log-rank $P > 0.5$), Fig. 2e-f.

2.2.4 The LM reproduces known lifespans of genetic perturbations

To test whether the LM could be used to characterize the effects of mutations and RNAi on lifespan, we divided wild type animals and *daf-16(mu86)* mutants3 between plates seeded with either an *E. coli* HT115(DE3) strain containing an RNAi construct targeting the insulin/IGF receptor *daf-2*3, 22 or an empty vector. As expected, *daf-16* mutants live shorter than wild type animals, and knockdown of daf-2 by RNAi extended the lifespan of wild type animals but not *daf-16* mutants (Fig. 3a). We found no significant difference between manual and automated

survival curves for all conditions (Log-rank $P > 0.3$) except for those of wild type animals feeding on control bacteria (Log-rank $P < 0.001$), which appeared to live 21 hours longer in the manual assay. A subsequent replicate showed no significant difference (data not shown). These experiments were performed at 25 °C; we observed corresponding effects at 20 °C (Supplementary Note 13).

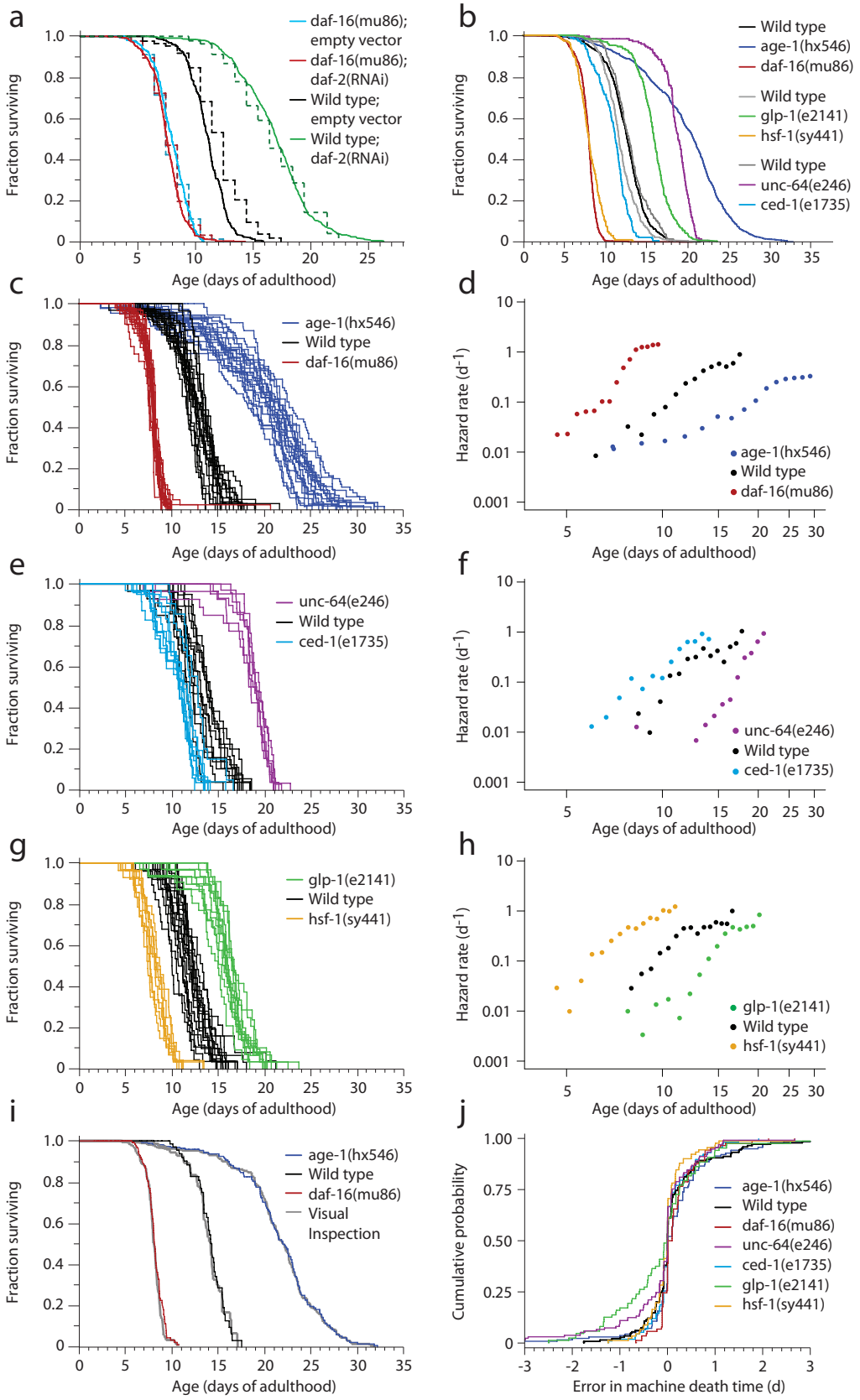

Figure 2.3: **Automated mutant survival data.** Caption is continued on the following page

Figure 2.3 (Cont'd): Automated mutant survival data. The manual experiments required one hour per measurement once a day, whereas the automated method required no experimenter time to collect twenty-four daily measurements of significantly larger populations. **(a)** Wild type animals and daf-16(mu86) mutants were fed bacteria containing either an empty vector or a vector for daf-2(RNAi). These populations were observed either by the automated (solid, 2,015 animals) or manual (dashed, 541 animals) method. **(b)** In addition, six mutant populations were monitored by the LM. Their lifespan data are shown both aggregated and, in subsequent panels, grouped by plate. **(c-h)** Each panel pair refers to a separate scanner also running a wild type population as reference. The population sizes were **(c,d)** daf-16(mu86): 654, wild type: 594, age-1(hx546): 1109. **(e,f)** ced-1(e1735): 255, wild type: 314, unc-64(e246): 193. **(g,h)** hsf-1(sy441):234, wildype: 472, glp-1(e2141): 335 **(d, f, h)** Hazard rates were estimated as in Fig. 2d (Supplementary Note 11). **(i)** As in Fig. 2e, a comparison between survival curves determined from visually and automatically scored death times validates the LM operation. Comparisons for all other mutants are shown in Supplementary Note 13. **(j)** The cumulative LM error (see Fig. 2f) for each mutant population provides an alternative view of validation. Experiments performed by BU, ZN, ILM, and NES.

Consistent with previous reports [Friedman and Johnson, 1986, Larsen et al., 1995, Ailion et al., 1999, Kenyon, 2010, Haskins et al., 2008], mutations in the phosphoinositide 3-kinase *age-1*, the syntaxin homolog *unc-64*, and the Notch receptor *glp-1* result in lifespan extension, whereas mutations in *daf-16*, the heat shock factor *hsf-1*, and the CD91 homolog *ced-1* shorten lifespan. Our automated technique correctly identifies known genetic determinants of *C. elegans* lifespan (Fig. 3b—h). Although population sizes (Supplementary Table 1) were too small to compare parametric models, we used these data to estimate27 the hazard rates of each mutant (Fig. 3d,f,h, Online Methods).

Visual image inspection (Fig. 3i,j) validates the LM-acquired survival curves of mutants. We note that the machine performed well on the movement-defective mutants *unc-64(e246)*, *unc-4(e120)*, *unc-50(e306)*, and *unc-119(ed3)* (Supplementary Note 13), even though the image analysis parameters were optimized for wild type movement. The analysis of *glp-1(e2141)* performed equally well as that of wild type, despite mutant animals being thinner from the absence of a germline.

2.2.5 The LM enables high-resolution stress resistance assays

Survival assays are widely used to evaluate the ability of individuals to withstand exogenous stresses. Exposure to high temperature (35 °C) or toxins dramatically shortens the lifespan of *C. elegans*. Mean survival ranges from several hours to a few days [Lithgow et al., 1995, Tullet et al., 2008], making data collection at high frequency challenging for the manual approach. We evaluated the performance of the LM in two stress-resistance scenarios: exposure to high temperature [Lithgow et al., 1995] and the oxidant tert-butyl hydroperoxide (t-BuOOH) [Tullet et al., 2008]. The LM required no modifications beyond adjustments to a subset of parameters used by our image analysis software to quantify worm movement and identify worm death times (Methods).

Age-synchronous wild type, *age-1(hx546)*, and *daf-16(mu86)* mutants were grown under standard conditions at 25 °C and shifted on the second day of adulthood to one of four conditions: agar plates at 25 °C or 35 °C and agar plates with 3 mM or 6 mM t-BuOOH at 25 °C. While no wild type animals died during the first two days on control plates at 25 °C, their lifespan was shortened dramatically at 35 °C. In agreement with previous reports [Lithgow et al., 1995], age-1(hx546) mutants lived longer than wild type and *daf-16(mu86)* mutants showed a small but statistically significant reduction in survival (Fig. 4a,b). The results obtained by visual and automated image analyses were highly correlated (Fig. 4c,d). We observed a similar, concentration-dependent effect of t-BuOOH on lifespan, consistent with previous studies28 (Fig. 4e,f). The LM is therefore well suited for assaying the effects of chemicals on survival in *C. elegans*.

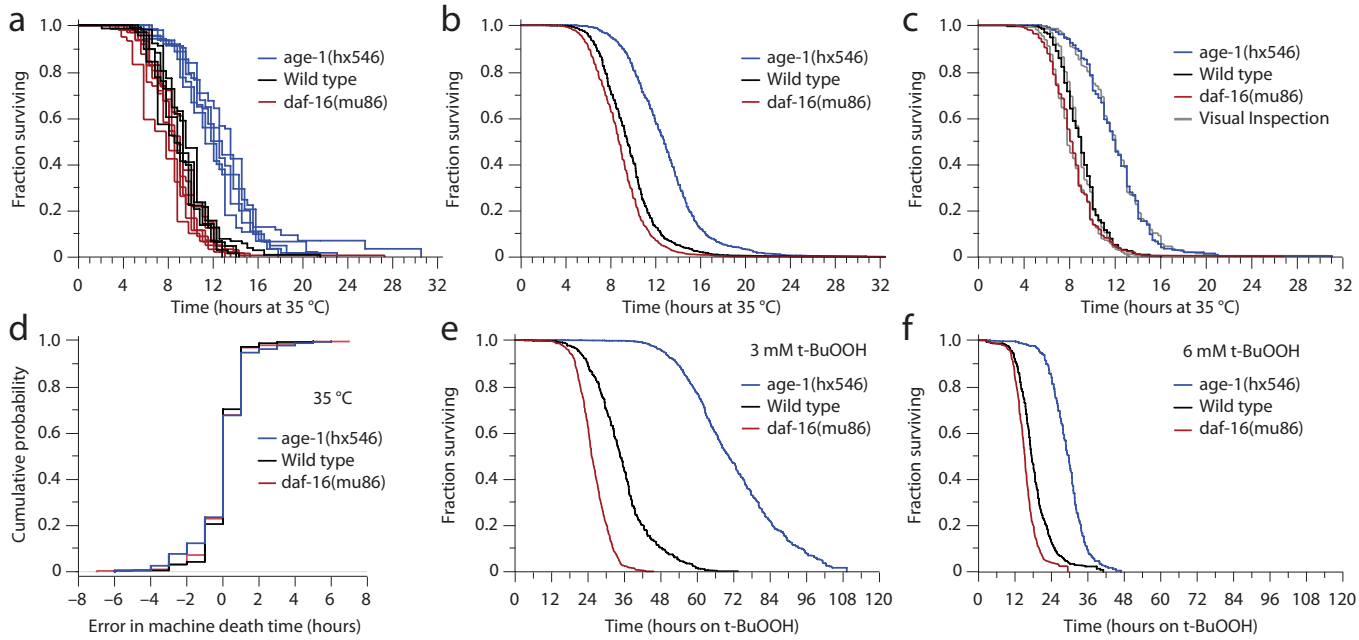

Figure 2.4: Automated stress-resistance assays.(a,b) Age-synchronous populations of 1,873 *age-1(hx546)* mutants, 2106 *daf-16(mu86)* mutants, and 1,726 wild type animals were transferred to 35 °C on their second day of adulthood and their death times were recorded by scanners operating at that temperature. Animals died over the course of thirty hours, as shown across the populations of individual plates **(a)**, and in aggregate **(b)** for all animals of each genotype. **(c,d)** The LM operation is validated as in Figs. 2e,f and 3i,j (Log-rank P > 0.5 in all cases). **(e,f)** LM-acquired survival curves are shown for populations of age-1(hx546), wild type, and daf-16(mu86) animals that were transferred onto **(e)** 3 mM t-BuOOH (2,135 animals) and **(f)** 6 mM t-BuOOH (1,808 animals total) at 25 °C. Experiments performed by BU and NES.

2.2.6 The hazard rate of animals under stress

We found the hazard rate function of wild type, *age-1(hx546)* and *daf-16(mu86)* at 35 °C to consist of an accelerating and decelerating phase (Fig. 5a), as at 25 °C, suggesting that the demographic characteristics of aging are preserved even under intense stress conditions. These features did not depend on the presence of live bacteria, as both were also observed in populations raised and assayed on UV-killed bacteria (Fig. 5b). To our knowledge, this is the first quantification of a hazard rate function for metazoans under extreme conditions.

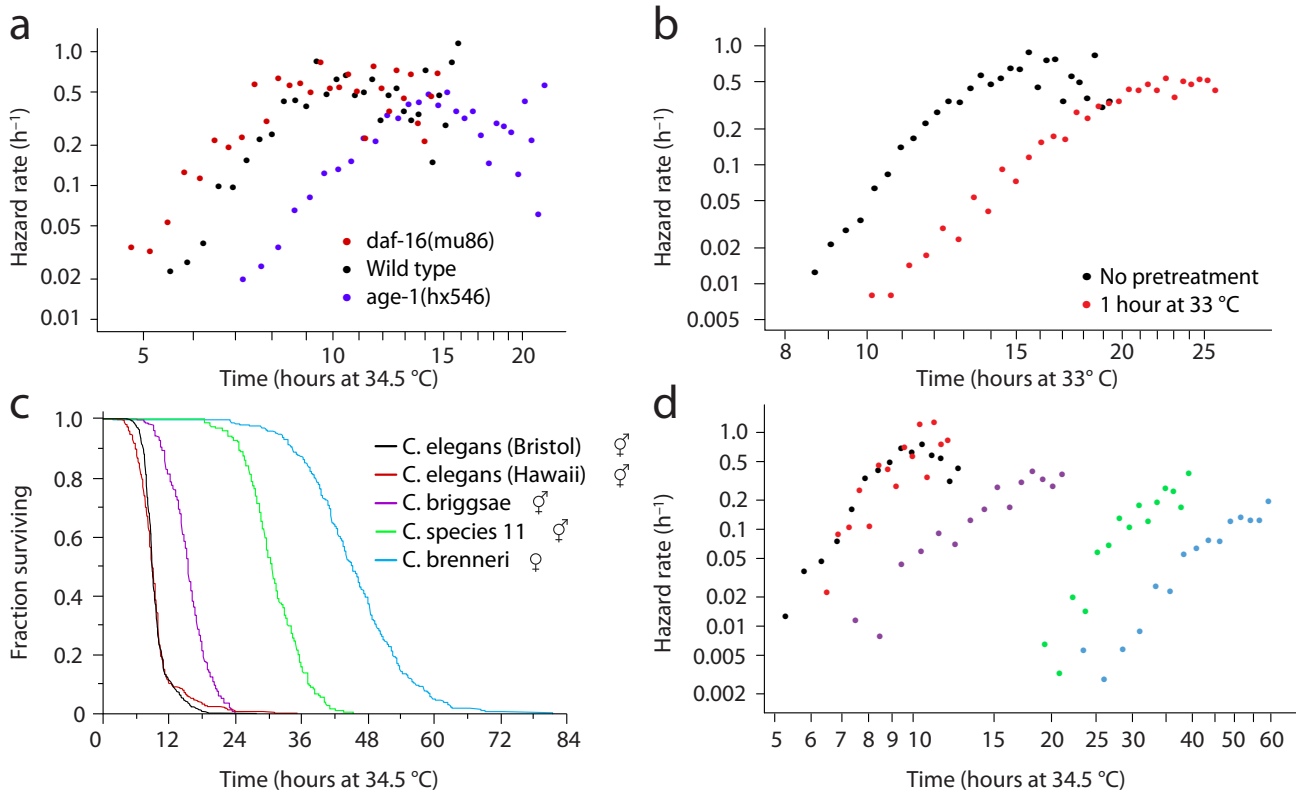

Figure 2.5: Hazard of animals under thermal stress.(a) Hazard rates *(dots)* were estimated as in Fig. 2d for populations of 1,873 age-1(hx546), 2,106 daf-16(mu86) and 1,726 wild type animals. **(b)** An acquired thermotolerance experiment is shown from a hazard perspective. Roughly 4,000 *C. elegans* were grown on UV-killed bacteria at 20 °CC. On the first day of adulthood, half of the animals were transferred to 33 °CC for one hour and allowed to recover for twelve hours at 20 °CC. Treated animals were then transferred a second time to 33 °CC and their survival determined with the LM. The hazard rate curve of 2,022 pre-treated animals was compared to that of 2,459 untreated animals. **(c)** The panel shows LM-acquired survival curves obtained from populations of 1,306 *C. elegans* (N2 Bristol isolate), 255 *C. elegans* (Hawaiian isolate), 285 *C. briggsae*, 224 *C. species 11*, and 295 *C. brenneri* nematodes raised on live OP50 at 20 °CC and transferred to 34.5 °CC on the second day of adulthood. **(d)** Hazard rate plots for the Caenorhabditis species of panel (c). Experiments performed by BU, ZN, ILM, and NES.

The effect of thermal stress appears to consist primarily in rescaling time, while barely impacting the coefficient of variation (CV) of the lifespan distribution (Supplementary Table 1). Up to median lifespan and barring frailty effects, 1 hour for animals at 25 °C corresponds to 1.8 minutes for animals at 35 °C and the latter experience at every moment a 33-fold higher risk of death. Likewise, the effect of *age-1(hx546)* and *daf-16(mu86)* mutations relative to each other at 35 °C

consists primarily in a change of time scale (Fig. 5a) with a marginal effect on the CV. This is a marked departure from the behavior at 25 °C (Fig. 3d) where DAF-16 increases and AGE-1 decreases lifespan variation disproportionately relative to the mean.

Previous studies in *C. elegans* have shown that a brief heat pulse induces a heterogeneous stress response, whose magnitude predicts survival upon a subsequent shift to high temperature [Lithgow et al., 1995, Rea et al., 2005]. We evaluated the effect of pretreatment on thermotolerance and found a substantial stretching of the time scale (Fig. 5b).

Despite a separation of 100 million years or more [Kiontke et al., 2011], the Caenorhabditis species *elegans*, *briggsae*, *brenneri*, and *species 11* are sufficiently similar in appearance for the LM to function correctly with no additional calibration at 35 °C. We confirmed previous reports [Amrit and May, 2010] that *C. briggsae*, and *C. brenneri* are more thermotolerant than *C. elegans*, and found this to be the case also for C. species 11. The thermotolerance profiles of these species correlate with the temperatures of their natural habitats [Kiontke et al., 2011], with tropical species being more thermotolerant than temperate ones. As we observed for *C. elegans* mutants, the genetic background of *Caenorhabditis* species, too, causes predominantly a rescaling of time, while having a much smaller effect on the coefficient of variation of the lifespan distribution (Supplementary Table 1).

2.3 Discussion

By using standard nematode culture conditions, our method meaningfully extends, within limits (Supplementary Note 16), the existing experimental literature and can be checked against it. The modular design of the LM gives researchers the flexibility to expand an installation by simply adding scanners (Supplementary Notes 14 and 15). A single scanner is sufficient to compare the lifespan of a mutant population against a wild type control, detecting 10% differences in survival with 99% confidence and 99% power. Facilities of ten or more scanners provide sufficient throughput to systematically characterize the effects of many genes and environmental conditions on survival. At even larger scale, scope can be paired with depth by evaluating samples that are large enough to resolve and quantitatively compare hazard rate features, thus linking

interventions at the molecular level with their statistical footprint at the organismic level.

2.4 Methods

***C. elegans* culture**. The following nematode strains were used: CB246: *unc-64(e246)*; CB3203: *ced-1(e1735*); CF1037: *daf-16(mu86)*; HT1593: *unc-119(ed3)*; N2: wild type (Bristol); QZ112: *hsf-1(sy441)*; QZ121: *glp-1(e2141)*; TJ1052: *age-1(hx546)*; TJ1060: *spe-9(hc88)I; fer-15(b26)*; TJ1062: *spe-9(hc88)I; fer-15(b26) age-1(hx542)*). Worms were cultured under standard conditions [Stiernagle, 2006] on 6 cm-diameter Petri dishes containing 12 mL of NGM agar (referred as "standard plates"). For survival assays, worms were housed on "assay plates": shallower, 5 cm-diameter plates (VWR 351006) containing 8 mL of NGM agar32 in which CaCl2 was omitted to prevent formation of an insoluble precipitate that reduces image quality, 22.5 μg/ml nystatin (Sigma N3503) was added to prevent fungal growth, and 10 μg/ml 5-fluoro-2'-deoxyuridine (FUDR, Sigma) was added to eliminate live progeny. Fungicide and FuDR are not strictly necessary but increase the throughput of the method.

Worms were grown at either 20 °C or 25 °C as noted and fed *E. coli* OP50, except in RNAi experiments where they were fed *E. coli* HT115 (DE3) bacterial strains. hsf-1 failed to develop at 25 °C and so were cultured at 20 °C and transferred to 25 °C at the L4 molt, along with their wild type control. glp-1 animals were transferred as eggs from 20 °C to 25 °C, along with their wild type control. The *glp-1* and *hsf-1* wild type controls did not differ and are aggregated in Fig. 4g,h.

We monitor the condition of all plates throughout in every experiment by examining plate images and videos. Plates that exhibit fungal invasion, contamination, desiccation, fogging, or appear otherwise compromised are excluded from analysis. On average 10% of plates are excluded in this way, primarily as the result of fogging or desiccation. The bacterial lawn remained present on the plate at the conclusion of each experiment and provided sufficient food for all animals on the plate. Seeding plates with more bacteria than what we normally use does not affect lifespan (data not shown). Some nematode strains used in this work were provided by the Caenorhabditis Genetics Center, which is funded by the NIH National Center for Research Resources (NCRR).

We aim to place roughly 35 animals on each plate under automated observation, which ensures

that sufficient bacteria are present to feed nematodes throughout each experiment, confirmed by visual inspection of bacterial lawns. Population sizes are not constrained by any step of image processing. When animals were deposited onto plates as a liquid suspension, a natural variation exists in the number of animals transferred per plate. For the wild type population shown in Fig. 2b, we found no correlation between population size and mean lifespan of each plate ($P = 0.67$). We also found no correlation between the population size and frequency of multi-worm clusters (Supplementary Note 8) observed on each plate ($P = .82$), suggesting that higher population densities are feasible, at least for wild type animals.

Age synchronization. For survival assays, age-synchronous cohorts were prepared by hypochlorite treatment [Wilkinson et al., 2012]. Eggs were washed in M9 buffer [Dong and Suen, 2005] 4-6 times and immediately placed on standard plates containing *E. coli*. Late-L4 larvae were transferred onto standard plates containing 10 μg/ml FUDR (established by visual inspection of vulval development). On the second day of adulthood, worms were transferred onto assay plates. Prior to being plated, wild type eggs used in Fig. 2 were subjected to an additional synchronization step by arresting their development at the L1 larval stage, accomplished by placing them in sterile M9 buffer with rocking for 14 hours at 25 °C.

Plate loading. In preparation for imaging, groups of sixteen assay plates containing worms were placed, top removed, facing down on a glass sheet. We seal these plates against the glass sheet in a 4 x 4 grid by means of a rubber mat (Supplementary Note 1). The use of this rubber gasket helps to prevent agar from desiccating for the duration of the experiment. When loading samples into scanners, glass sheets act as trays allowing easy, fast experimental setup.

Plate drying. The plates on which animals are housed during an automated lifespan assay must be prepared carefully to ensure that the height of the agar matches the modified focal plane of a scanner. These plates are poured using a peristaltic pump (Wheaton Unispense) and then dried carefully to reach a consistent osmolarity and agar height. Drying was performed by sealing 75 plates inside a plastic bin containing 942 g of desiccated $CaSO_4$ (drierite), which, in our experience, absorbs roughly 90 g H_2O / kg. After 12 hours at room temperature, the sealed bin reaches equilibrium, allowing precise, reproducible plate drying. The weight of drierite used must

be monitored and adjusted over time, as drierite appears to gradually become less absorptive. The dimensions of the bin do not appear to play a significant role in the drying process. To reach a final plate osmolality equivalent to the stock recipe, we add an extra 5% of H2O to our initial mix and calibrate the drierite mass such that 5% of the each plate's weight is removed during drying. Scanner focus seems fairly robust to variations in final plate volume; our experiments were performed on plates varying in final weight between 13.2 g and 14.2 g. After drying, plates can be sealed and stored at 4 °C until needed. Plates can be seeded with *E. coli* any time after drying.

Temperature calibration of scanners. Scanner temperature was measured by mounting a thermocouple (ThermoWorks USB-REF) on the bottom of an empty Petri dish, and loading that dish onto the center of each scanner. Glass sheets and rubber mats were used as described above. Temperature variability is minimized by moving scanner power sources outside the incubator, and by positioning scanners in a staggered vertical orientation to facilitate airflow throughout the incubator (Supplementary Note 1). Scanners produce their own heat and are warmer than ambient incubator temperatures. The temperature gradient between glass and plates prevents condensation from forming on the glass. To increase the robustness of this effect, we add the anti-fog coating Rain-X to the surface of the glass trays.

Scanning schedule. Scanning schedules reflect a compromise between heat production, image resolution, and scanning speed. Images captured at high resolutions allow nematodes to be distinguished with greater clarity, and a high frequency of capture allows their death times to be identified more precisely. However, high image resolution requires a slower scanning speed, yielding distorted images of moving animals, while high capture frequencies allow scanners less time to cool between scans. Therefore, we limit our captures to 3200 dpi images every 15 minutes. An additional complication arises in that scanners operate much more slowly when imaging wide areas than they do imaging narrow strips, perhaps due to a processing limitation of scanner electronics. To balance these factors, we collect images of one column (consisting of 4 plates) at a time rather than all columns at once. In this mode, capture is complete in about 8 minutes, indicating that the scanner bar moves at an average rate of 400 microns per second. We image each column twice in succession, with the first column being scanned at t and $t + 15'$; the second at $t + 30'$ and $t + 45'$; the third at $t + 60'$ and $t + 75'$; and the fourth at $t + 90'$ and $t + 105'$.

The first column is then scanned again at $t + 120'$, and so on. Hence, every column of four plates is imaged twice in a 2-hour cycle, with one short (15') and one long (105') time interval between successive images of any given plate, providing two timescales for movement detection.

Worm identification. Scanners capture images at a lower frequency and lower quality than attainable with conventional camera-mounted microscopes and thus require more de-noising and image registration (Supplementary Note 2). Based solely on image data, the LM must distinguish worms from other objects present on the plate, such as bacterial clumps and trail shadows. To discriminate between worms and non-worms, we use a Support Vector Machine (SVM)33 that classifies foreground objects based on a set of morphological features (Fig. 1; Online Methods and Supplementary Note 2). The SVM was trained once with a hand-annotated set of images of worm and non-worm objects. The resultant worm classifier appears to be general, yielding accurate results for young and old animals across strains.

Death determination. Manual lifespan assays declare death when an animal fails for the first time to respond to acute stimuli soliciting motion, after which the animal is discarded. The LM does not provide such mechanical stimuli. Instead, it runs a retrospective analysis of posture change to determine the cessation of all spontaneous (as distinct from stimulated) motion, as described in the main text and detailed in Supplementary Note 5.

Censoring. An animal may be lost before its death from age is observed; for example, it may get stuck on a plate wall outside the field of view. Such losses cannot be ignored, and must be censored to obtain proper survival statistics (Supplementary Note 8). Our image analysis maintains a tally of worms at any given scan and uses this to infer worm losses. This process is vitiated when worms aggregate into clusters, inside which our SVM often cannot accurately classify them (Supplementary Note 2). After leaving a cluster, worms are again recognized and accounted for. This temporary absence, however, makes it impossible to unambiguously reconstruct the exact timing at which true losses occur. It is nonetheless possible to make reasonable estimates of their timing, as discussed in Supplementary Note 8. When two (or more) worms die within a cluster, their death times can be estimated through rapid inspection and annotation of the visual records using the Worm Browser (Supplementary Video 6 and Note 4).

Hardware assembly. Please refer to the detailed assembly instructions in Supplementary Note 3.

Imaging servers and image-processing servers. Ten scanners (Epson v700) are housed in each incubator (Thermo Forma 3920). Each group of twenty scanners is controlled by a single Linux server, running on consumer PC desktop hardware. Communication is accomplished over USB cables using modified SANE scanner drivers. The servers store image capture schedules in a mySQL database and execute captures in parallel across multiple scanners. Captured images are buffered to a local disk drive, and asynchronously copied to a network storage server. A system that scans every 15 minutes over the course of a typical 28-day lifespan experiment generates about 100 GB of compressed image data per scanner (16 plates). Processing intermediates require an additional 100 GB, after which all imagery can be deleted. A central linux command-and-control server hosts a web interface through which experimental metadata is specified, including scanning schedules and image processing tasks.

Data inspection and validation software. A cross-platform graphical user interface (GUI) client was developed (the "Worm Browser") to facilitate the visual validation and curation of image data banks. Images of each animal can be viewed in aggregate and sorted by their lifespan. Time-lapse videos of individual worm deaths can be annotated. The Worm Browser integrates both automated image analysis results and manual annotations to produce a variety of statistical outputs. Outputs are formatted as comma separated value (CSV) files for processing by separate statistical packages, including R, JMP, and SAS.

Automated survival assay setup times. The time required for one person to complete a 6,000 worm lifespan experiment (as in Fig. 2) is as follows. Plate pouring: 3 h; plate drying 2 h; worm synchronization and transfer to plates with FUdR: 7 h; transfer to assay plates: 4 h; loading plates into scanners and scheduling scans: 1 h; performing 25 °C assay: 0 h (user time) / 28 d (chronological time); image analysis and validation: 2.5 h (user) /8 d (computer time on one quad-core i7).

Manual lifespan assays. Animals were synchronized and grown at 25 °C as described above, and inspected daily for response to prodding under a dissecting microscope, according to

standard practice6, 7. Dead animals were removed upon identification.

RNA interference. We performed RNAi by feeding as described [Dillin et al., 2002] on plates containing 100 μM ampicillin and 400 μM isopropyl b-D-thiogalactopyranoside (IPTG). RNAi bacteria contained either an empty vector (pAD12) or one targeting daf-2 (pAD48)24.

UV-treatment. For assays performed on UV-treated NEC937 (*E. coli* OP50 ΔuvrA::KanR) bacteria14, killing was accomplished by irradiating freshly seeded plates with 1 J/m2 of 254 nm light in a UV Stratalinker (Stratagene). Plate sterility was further ensured through addition of 100 μM ampicillin.

Automated stress survival assays. Prior to exposure to 35 °C or t-BuOOH, animals were synchronized and grown at 25 °C as described above. Automated survival assays were performed at 35 °C on animals on their second day of adulthood. For t-BuOOH survival assays, plates were prepared by adding t-BuOOH (Sigma) to a final concentration of 3 mM or 6mM into molten agar, twelve hours before the start of the assay. All plates were seeded with 5x concentrated *E. coli* OP50, washed 5 times in M9 buffer. Automated t-BuOOH survival assays were performed at 25 °C on animals transferred onto assay plates containing t-BuOOH on their second day of adulthood. In both 35 °C and t-BuOOH survival assays, the assay plates did not contain nystatin or FUdR. In the experiment involving multiple Caenorhabditis species, animals were grown on live OP50 bacteria at 20°C and transferred to 35 C on the second day of adulthood.

To enable automated analysis of deaths during exposure to high-temperature and t-BuOOH, we adjusted two image-processing parameters relative to un-stressed lifespan experiments. First, we reduce the minimum duration of time that animals must cease crawling before being marked “stationary”. Second, we reduce the minimum duration of time that animals must be completely motionless before being marked “dead”. The movement score threshold used as the criteria for absence of motion remains identical for stress resistance and lifespan assays.

Statistical methods. Log-rank and Wilcoxon tests for homogeneity between two or more survival curves were performed using JMP (SAS) [Sall et al., 2012] and R [R Development Core Team, 2009]. Categorical regression using the Buckley-James estimator [Buckley and James, 1979] was performed in R using the surv and rms packages. Estimates of the piecewise-constant

hazard rate were made in R using the muhaz package. Weibull and Gompertz parameters were estimated by parametric regression in R using the flexsurv package.

Chapter 3

Wild-type *C. elegans* exhibit a Weibull distribution of death times at 25 °C

3.1 Parametric Models of Survival Data

The distribution of lifespans contains the statistical signature of underlying stochastic mechanisms that act to determine the timing of death. Nearly two hundred years ago, the British mathematician Benjamin Gompertz observed that the likelihood of a person's death rose geometrically in consecutive decades [Gompertz, 1825], suggesting that the statistical dynamics might be well approximated by simple mathematical formula. This raised the possibility that some fundamental principle might exist underlying aging to produce such a simple pattern. Gompertz suggested that such a principle might be made comprehensible and discussed by representing mortality data by parametrized mathematical functions. The modern framework for discussing death times involves three interrelated quantities, the survival function, $S(t)$, the lifespan distribution $f(t)$, and the hazard function, $h(t)$.

The survival function is defined simply as the probability that a death event, represented by the random variable T, has not yet occurred at time t; which is to say that the value of T is larger than t:

$$S(t) = \text{Prob}\{T > t\}. \tag{3.1}$$

3.2 Lifespan quantifies the outcome of complex aging processes.

Note: Select portions of this text are taken from the manuscript The *C. elegans* Lifespan Machine, presented in its entirety in chapter 2.

In this way, survival is defined as a cumulative distribution, the probability that an individual is still alive at time t. In experimental practice, survival functions are most often estimated directly from empirical data using the product-limit framework formulated by Kaplan-Meier [Kaplan and Meier, 1958a]. However, in circumstances when prior information is available about the statistical form of the distribution of death times in a population, the survival function can be specified in terms of a parametrized model, $S(t, \alpha, \beta) = \text{Prob}\{T > t|\alpha, \beta\}$. Parameter values can then be estimated from the empirical data, using a variety of techniques including (as done pervasively in this document) maximum likelihood estimation [Singer and Willett, 2003, Jackson et al., 2010]. An important property of a well-chosen parametric model is that the parametric $S(t)$ will closely match the non-parametric Kaplan-Meier $S(t)$.

The probability of death at a given time t, usually referred to as $f(t)$, is defined as $\text{Prob}\{t < T < t + dt\}$, or in relation to the corresponding survival function

$$f(t) = -\frac{dS(t)}{dt} \tag{3.2}$$

This allows us to define a very interesting and fundamental quantity, the hazard function, $h(t)$. The hazard function is the *conditional* probability of dying in the next instant, given that an individual has survived to time t, that is, $\text{Prob}\{t < T < t + dt \,|\, T > t\}$. Note the difference to the lifespan distribution $f(t)$. Specifically, $f(t)dt$ is the probability of dying within the time interval $[t, t + dt]$ calculated at age 0; whereas $h(t)dt$ is the probability within $[t, t + dt]$ calculated at age t and is thus (ideally) an instantaneous rate. The hazard function is defined as

$$h(t) = \frac{f(t)}{S(t)} \tag{3.3}$$

and specifies the rate by which animals die. Mair et al. [Mair et al., 2003] demonstrated that the

hazard rate of *melanogaster* can be instantaneously altered living animals by removing food or shifting the animals to a different temperature, suggesting that the hazard rate directly summarizes an important process acting throughout life to determine mortality. Dillin at al. [Dillin et al., 2002] demonstrated that genetic perturbations to *C. elegans* mid-life could alter a population's remaining lifespan, similarly demonstrating that the ongoing activity of molecular components in adults act to specify a population's hazard rate. The likelihood of an animal death appears to be determined by some ongoing molecular process or processes that act stochastically to determine whether or not animals die in the near future. The hazard rate can be used as a formal representation of this likelihood, defined as

$$\mathcal{L}(t, h(t, \theta)) = h(t, \theta) \tag{3.4}$$

This means that, given a model of a population's hazard rate, an individual's likelihood of dying at age t is specified by that model and its parameters. The molecular determinants of death acting within a time interval $[t_1, t_2]$ can be studied via their effect on an animal's likelihood of death in that interval, defined as $\mathcal{L}([t_1, t_2], h(t, \theta)) = \int_{t_1}^{t_2} h(t)dt$.

This definition matches our intuition–a group of old animals should experience a greater rate of death than a population of young animals, and this is summarized by the old group having a higher hazard rate.

Another commonly used quantity is the cumulative hazard $H(t)$, $H(t) = \int_0^t h(\tau)d\tau$, which can be used to express the survival function $S(t) = exp(-H(t))$, which results from integration of equation (S11) written as $h(t)dt = -d\log(S(t))$.

Like the survival function, the hazard function of a population can be estimated in an unbiased way through a variety of techniques [Hess et al., 1999, Mueller and Wang, 1994, Wang et al., 1998], which, however, all require much larger population samples than the estimation of survival, as the slope of the logarithm of the survival curve must be estimated precisely within small time intervals. Experimental acquisition of the data required to estimating hazard functions is technically challenging and time-consuming, limiting its exploration to just a few

studies [Curtsinger et al., 1992, Carey et al., 1995, Vaupel et al., 1998, Vanfleteren et al., 1998, Johnson et al., 2001, Mair et al., 2003, Baeriswyl et al., 2009, Wu et al., 2009].

Many attempts have been made to evaluate the underlying functional form of the hazard. This serves a practical purpose, as parameterized models provide strong assumptions that can be used to increase the power of various statistical tests [Singer and Willett, 2003], allowing conclusions to be drawn from smaller populations. Parametric modelling of the hazard rate also serves a conceptual purpose, in that the functional form of the hazard distribution derives from the stochastic dynamics of underlying mechanisms. By comparing unbiased estimations of the hazard rate to parametric models, theories about those stochastic dynamics can be tested against empirical observation.

Many parametric models have been used to describe death times, but here we address three models widely observed in failing systems.

$$h(t) = c \qquad \text{(Constant hazard)} \tag{3.5}$$

$$h(t) = ae^{bt} \qquad \text{(Gompertz hazard)} \tag{3.6}$$

$$h(t) = \frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1} \qquad \text{(Weibull hazard)} \tag{3.7}$$

The constant hazard model defines an individual's risk of death as unchanging over time. Individuals exhibiting a constant hazard rate are not aging, since the conditional probability $\text{Prob}\{t < T < t + dt \,|\, T > t\}$ does not depend on t; i.e. such individuals are no more likely to die as time progresses. Radioactive particles exhibit this behavior, as inferred from radioactive decay. Populations of radio-nuclides dwindles according to $S(t) = \exp(-ct)$ with half-life $(\log 2)/c$, since $S(t) = \exp(-\int_0^t h(\tau)d\tau) = \exp(-ct)$.

The Gompertz model formalizes Gompertz's empirical observation of geometrically increasing hazard rates in human populations. Gompertzian models are increasingly used in *C. elegans* genetics to classify purturbations relative to their effect on the model parameters [Samuelson et al., 2007]. One diagnostic characteristic of Gompertzian populations is that the logarithm of the hazard rate is linear on a semi-log (log-linear) scale, $\log(h(t)) = \log(a) + bt$. The survival

function of a Gompertzian population is $S(t) = \exp(a/b(1 - \exp(bt))$.

The Weibull model states that the hazard rate increases as a power of time, and has been shown to be a useful model for the failure times of a variety of mechanical systems [Rinne, 2008, Clark, 1991]. The Weibull hazard is a straight line on a log-log scale, $\log(h(t)) = \log(\alpha/\beta^{\alpha}) + (\alpha - 1)\log(t)$. This particular parametrization is chosen such that the survival function takes a clean form, $S(t) = \exp(-(t/\beta)^{\alpha})$, which is a so-called stretched exponential.

3.3 Comparing distributions of death times

3.3.1 General approach

There are many ways to ask whether a parametric model is a good representation of empirical data. The most common approach, used throughout this document, goes as follows. 1) Assume, to start, the model is a perfect match of the distribution underlying the experimental data 2) identify parameter values for the model most consistent with the experimental data. 3) Using this "best" parameter set, compare some aspect of the model's behaviour (i.e. the shape of the survival curve, the hazard function, etc.) with that observed in the experimental data. This third step can be done quantitatively, for example calculating the Wilson residuals, as in [Vanfleteren et al., 1998], or using the AIC criterion [Burnham and Anderson, 2004, Singer and Willett, 2003]. When comparing models with the same number of parameters, the AIC criterion essentially summarizes the likelihood of the maximum likelihood parameter set of the model for the experimental data, allowing models to be compared using the same underlying techniques and formalism as is used to identify model parameters. Quantitative techniques are useful because they simplify model comparison down to a single number, but it is often more informative to investigate *how* a model deviates from experimental data than *how much* it deviates. Qualitative techniques based around graphical comparisons of a model and parameters work well for this.

3.3.2 Quantile-Quantile and Quantile-Residual Plots

One familiar graphical means for evaluating simply plots an empirical measurement, for example the Kaplan-Meier (KM) survival curve of piecewise-exponential hazard rate estimate, against the models prediction. This is done, for example, in Fig. 3.2d which overlays various models' predicted survival curve against the experimental KM estimate. The same is done in Fig. 3.2e, comparing model and empirical hazard functions. These serve to highlight gross deviations between data and model, but an even more sensitive representation is used extensively here. The Quantile-Quantile (Q-Q) plot will be familiar to demographers [Wilk and Gnanadesikan, 1968]. Essentially, a survival curve (and corresponding cumulative distribution of failure times defines a relationship between the fraction of animals alive and time (eq. 3.1). The inverse of the cumulative distribution of failure times is called the quantile function

$$Q(p) = F(t)^{-1} \tag{3.8}$$

The quantile function of a parametrized model is continuous. By inverting a Weibullian cumulative distribution of failure times, $F(t) = p = 1 - \exp(-(t/\beta)^{\alpha})$, one obtains the Weibull quantile function, which specifies the first time t at which the fraction q of individuals will have died.

$$Q(p) = \beta(-\log(1-q))^{\frac{1}{\alpha}} \tag{3.9}$$

The Q-Q plot relates the times $Q_1(p)$ and $Q_2(p)$ across the range of quantiles $0 \leq p \leq 1$. This is equivalent to plotting the relationship $F_1(t_1) = F_2(t_2)$, with t_1 on the abscissa and t_2 on the ordinate:

$$t_2 = Q_2(F_1(t_1)) \tag{3.10}$$

. The abscissa of a Q-Q plot represents the time at which curve 1 reached each quantile; the ordinate represents the time at which curve 2 reached the same quantile. Equivalent models indicate that $F_1(t_1) == F_2(t_2)$, and consequently eq. 3.10 resolves to $t_2 = t_1$. The quantile quantile plot will be a straight line with a slope of 1, intersecting the abscissa at $t = 0$.

For a Kaplan-Meier estimate of survival $S[t]$, the quantile function is represented by a set of discrete quantile-time pairs $(t_1, 1 - S[t_1]), (t_2, 1 - S[t_2]), ..., (t_N, 1 - S[t_N])$, where N is the number of unique times at which deaths were observed to occur. A parametric model can be compared to empirical data by evaluating the parametric $Q_{\text{parametric}}(p)$ at each KM pair $(t_i, F_{km}[t_i])$ [Rinne, 2008]. A model that perfectly predicts the empirical data will have the property that $Q_{\text{parametric}}(F_{km}[t_i]) == t_i$ for all i. Thus, a plot of $Q_{\text{parametric}}(F[t_i])$ vs t_i will closely approximate a straight line with a slope of 1, intersecting the abscissa at t=0.

Q-Q plots provide a very easy way for comparing Weibull distributions. Consider two Weibull distributions with the parameters (α_1, β_1) and (α_2, β_2). We want to plot the relationship $t_2 = Q_2(F_1(t_1))$. For Weibull distributions, this relationship can be expressed as

$$\log t_2 = \frac{\alpha_1}{\alpha_2} \log t_1 + \log \frac{\beta_2}{\beta_1^{\frac{\alpha_1}{\alpha_2}}} \tag{3.11}$$

Thus, on a Q-Q plot with log-log axes, two Weibull distributions will always form a straight line. The slope of this line will equal the ratio of the two distribution's α parameters. The line will intercept the abscissa at the time $\log \frac{\beta_2}{\beta_1^{\frac{\alpha_1}{\alpha_2}}}$, which for curves with the same α is clearly $\log \frac{\beta_2}{\beta_1}$. We use this property of Weibull distributions to make conclusions on the relationship between experimental replicates in chapter 3 and the relationship between lifespan at different temperatures in chapter 4.

Q-Q are fairly standard, but they are often difficult to interpret for the following reason. Establishing the relationship between two curves–to see if they are identical for example, requires comparisons between $t_2 = mt_1$ and $t_2 = (m + \Delta m)t_1$. This is surprisingly difficult around the point $m = 1$, but much easier around the point $m = 0$. At $m = 0$, even very small values Δm can be seen as subtle deviations from horizontality. Thus, in this document, data is presented as Quantile-Residual (Q-R) plots, which graph the relationship

$$.r = Q_2(F_1(t_1))/t_1 \tag{3.12}$$

The normalization factor t_1 has the effect of rotating the Q-Q plot by forty-five degrees clockwise.

Identical curves will produce Q-R plots that are horizontal lines at $r = 1$. Log-log Q-R plots comparing two Weibull distributions will form a straight line with slope $\frac{\alpha_1}{\alpha_2} - 1$. In this context, two curves where $\alpha_1 = \alpha_2$ will form a horizontal line a ratio $r = \frac{\beta_2}{\beta_1}$ which is very easily recognized by eye. An example of several Q-R plots is provided in fig 3.1

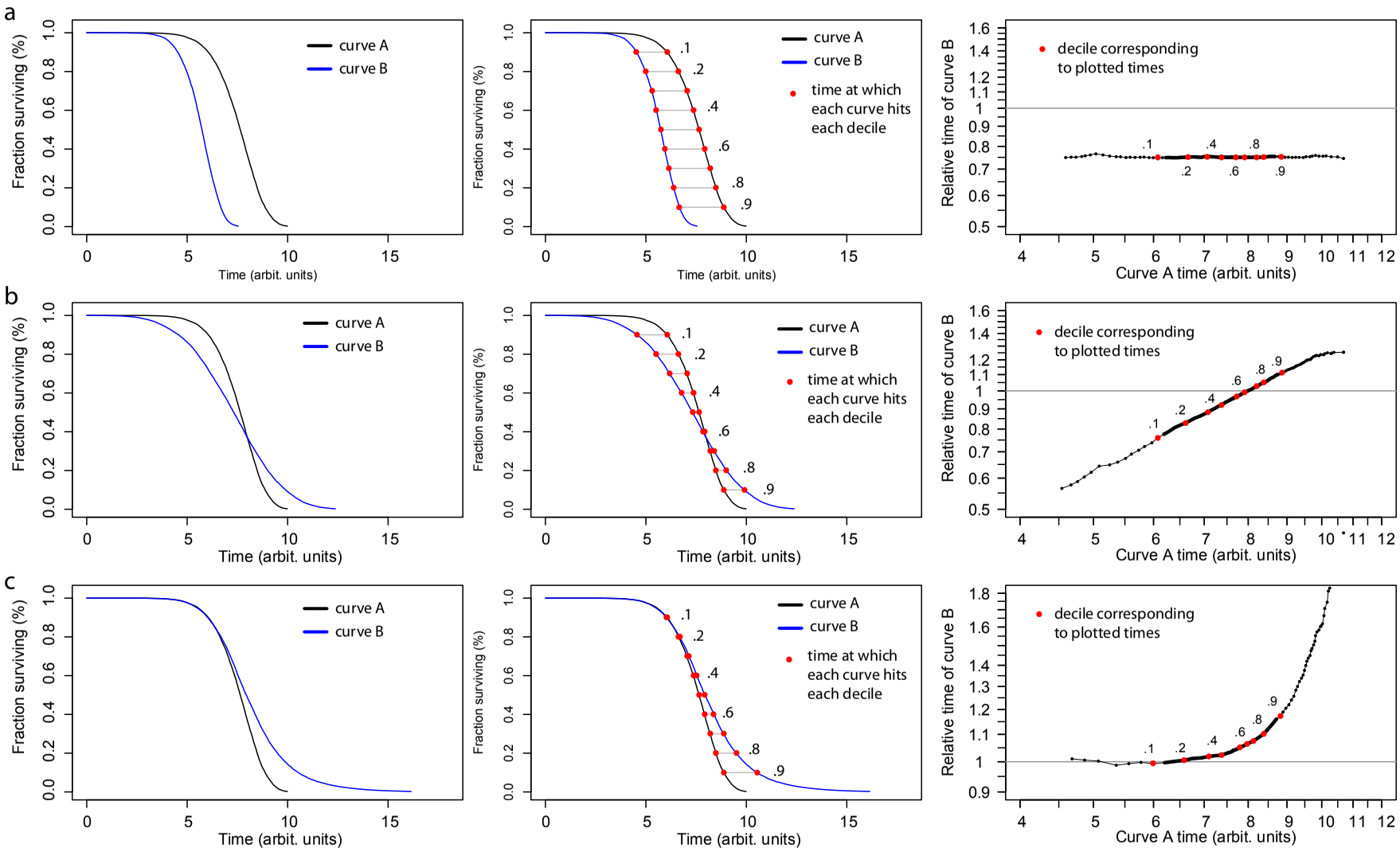

Figure 3.1: An explanation of Quantile-Residual (Q-R) plots. Caption continued on the following page.

Figure 3.1 (Cont'd): An explanation of Quantile-Residual (Q-R) plots. Two survival curves may differ in several ways (left), for example by **(a)** a simple rescaling (stretching) of time, or **(b)** by differing coefficients of variation. Two curves might even demonstrate **(c)** entirely different underlying statistical forms. These relationships can be explored using Q-R plots, which map the relative time that one curve reaches a certain quantile (fraction of animals dying) relative to another. These times are shown *(middle)* as horizontal lines connecting two survival curves. The time at which a reference curve *(black)* reaches each quantile is plotted on the abscissa (x-axis), with the relative time-to-quantile of a second curve *(blue)* is plotted on ordinate (y-axis). A horizontal line at $r = 1$ indicates two survival curves are identical. Horizontal lines at $r < 1$ indicate that the second curve is living shorter than the reference, while retaining a proportional ratio of survival (indicating the coefficient of variation remains constant between the two curves). Deviations from horizontality (b), or even linearity (c), allow qualitative statements to be made on the nature of perturbations to survival.

3.4 A Closer Look at the Wild-type *C. elegans* Hazard Function

Chapter 2 includes an experiment in which a survival curve and hazard function was obtained for 3,500 wild-type *C. elegans* animals (Fig. 2). This experiment agreed with previous reports [Vaupel et al., 1998, Johnson et al., 2001, Baeriswyl et al., 2009] that the hazard rate of wild-type animals exhibits a period of rapid increase, followed by a marked deceleration among the longest-living cohorts. Deaths occurring during the period of initial rapid increase in hazard (the first 50% of deaths) were fit by Gompertz and Weibull models. This early analysis showed that both Gompertz and Weibull models fit the data, with Weibull fitting the data slightly better among the earliest dying animals.

This experiment involved an unusually large population relative to most *C. elegans* studies–the largest population comfortably measured in a single incubator filled with flatbed scanners. However, the the population size was not sufficient to definitively demonstrate a Weibullian hazard function. Furthermore, the experiment constituted essentially one independant replicate. To get a closer look at the statistical form of *C. elegans* lifespan, and gain greater confidence in the conclusions of Chapter 2, a larger data needs to be compiled. Over the course of a year, the Fontana lab obtained lifespan distributions of approximately one hundred mutants at 25 °C. Analysis of these mutant curves is only in its earliest stages, but crucially for our purposes here, each of these experiments contained a large number of wild-type individuals included as control

groups for the mutants. Ten of these experiments included control populations of more than one thousand age-synchronous wild-type animals. These wild-type controls represented, in effect, ten independent replicates of the experiment presented in Fig. 3.2.

Across the ten experiments, median lifespan varied between 11.9 and 14.2 days (Fig. 3.1a, Table 3.1), suggesting that wild-type lifespan varies even when measured on standardized equipment. These significant (log-rank $p < .001$) differences indicate that some aspect of the worm handling, assay environment, lineage effect varies between replicates, with candidates including temperature, bacterial composition, plate osmolality, plate pH, or inherited epigenetic determinants resulting from prior starvation. Such effects persisted despite deliberate attempts to limit variability in these factors, including carefully monitoring temperature using thermocouples, development of a quantitative plate drying technique (2.4), and ensuring the parents of animals assayed were preceded by several generations of well-fed continuous growth.

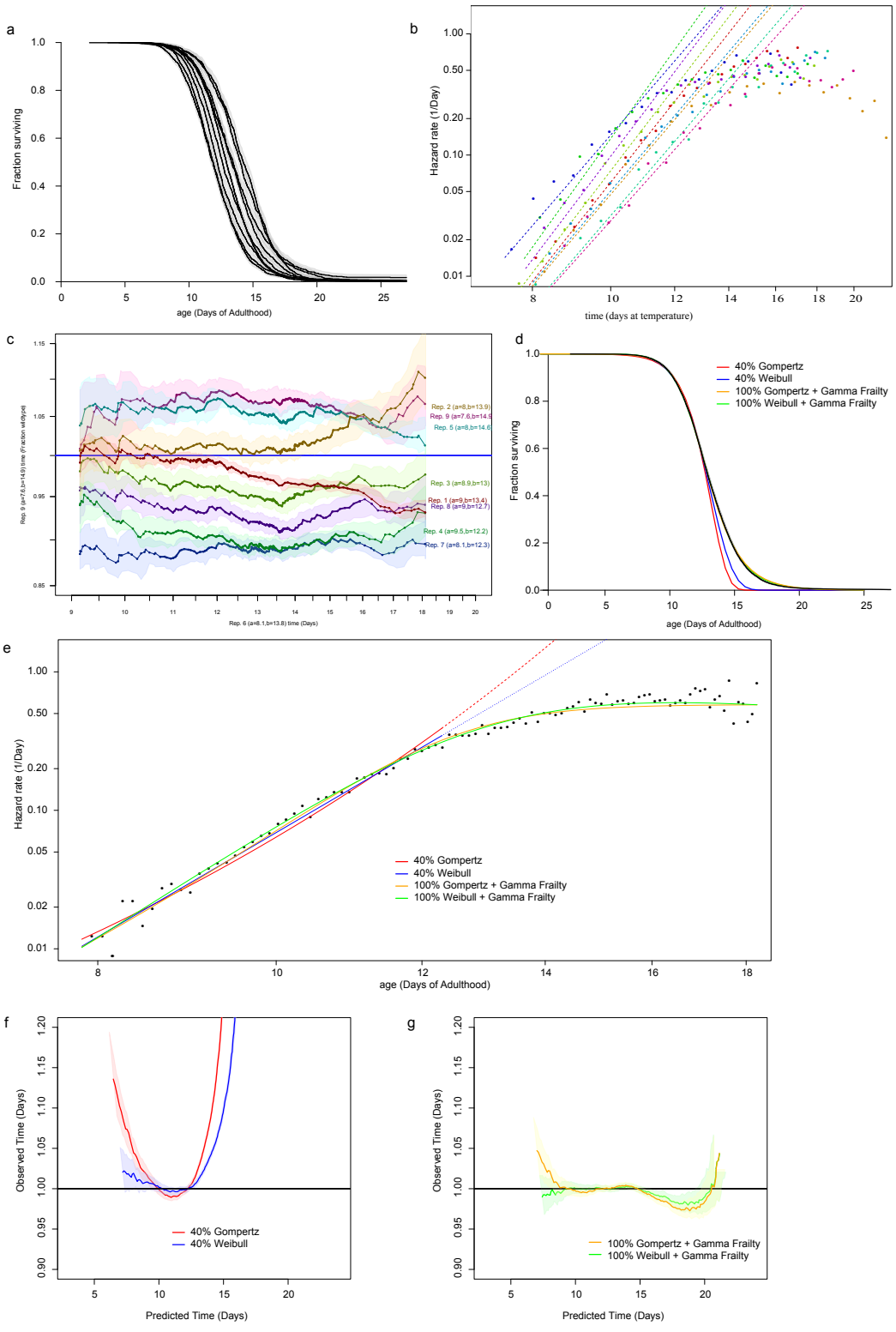

Figure 3.2: Wild type mortality at 25°C. Caption is continued on the following page

Figure 3.2 (Cont'd): Wild type mortality at 25°C. (a) The survival curves for each of 10 independent replicates of wild-type C. elegans, cultured on OP50-seeded solid agar, and monitored using the lifespan machine. Each curve contains more than one thousand individuals. **(b)**The same data was used to generate estimates of each population's hazard function, and ML estimates of the Weibull parameters based on the earliest 40% of deaths. **(c)** A quantile-residual plot uses replicate 6 as a reference to compare the shape of survival curves between replicates. For a description of quantile-residual plots, refer to 3.3.2. **(d)** Deaths across all replicates were aggregated using the AFT-model described in chapter 2, to produce a single 12,932 animal curve. MLE parametric distributions (colored lines) are plotted over the Kaplan-Meier estimated empirical survival (black line). **(e)** The hazard function was estimated for the large population, with MLE parametric distributions overlaid. Quantile-residual plots are shown, comparing the aggregate curve to **(f)** MLE Weibull and Gompertz estimates and **(g)** MLE Weibull and Gompertz estimates incorporating a Gamma-distributed frailty term. Experiments were performed by BU,ZN,ILM, and NES

Table 3.1: Summary data for ten replicates measuring the lifespan of wild type *C. elegans* at 25 °C. Nine replicates collected over two years were better fit by Weibull than Gompertz models in all but one case (Replicate 5). The addition of a Gamma-distributed frailty term did not alter the effect–a frailty model with a Weibullian baseline hazard fit the data better than Gompertzian baseline hazard in all but one replicate (replicate 9). In the aggregate curve formed from all replicates, the Weibull and Weibull + Gamma Frailty models fit extraordinarily better than the Gompertz.

Group	Deaths	Censored	Median Lifespan	Delta AIC Gompertz v. Weibull	Delta AIC Gompertz Frailty v. Weibull Frailty
Replicate 1	3945	607	12.92	26.82	18.50
Replicate 2	943	216	13.44	7.65	38.97
Replicate 3	1083	408	12.61	3.20	5.77
Replicate 4	1393	264	11.89	18.97	7.98
Replicate 5	1088	277	14.02	-1.28	3143.75
Replicate 6	917	190	13.32	12.49	444.54
Replicate 7	1468	36	11.88	19.66	31.21
Replicate 8	1218	194	12.21	4.58	782.52
Replicate 9	877	143	14.24	5.16	-1.25
All Experiment Aggregate	**12932**	**2335**	**12.88**	**104.65**	**127.97**

Table 3.2: Table 3.2: MLE Parameters at 25 °C. The maximum-likelihood estimate model parameters are provided for each model fitted, with 95% confidence intervals shown.

	Weibull Shape	Weibull Scale	Gompertz Rate (x103ˆ)	Gompertz Shape	Weibull Frailty Shape	Weibull Frailty Scale	Weibull Frailty Theta	Gompertz Frailty Rate (x103ˆ)	Gompertz Frailty Shape	Gompertz Frailty Theta
R 1	9±0.8	13.4±0.2	0.016±0.014	0.82±0.07	9.5±0.8	13.2±0.2	0.6±0.2	0.008±0.008	0.9±0.09	0.95±0.12
R 2	8±1.5	13.9±0.4	0.039±0.065	0.71±0.13	9.9±1.9	13.1±0.4	1.5±0.5	0.003±0.008	1.01±0.22	1.13±0.45
R 3	8.9±1.5	13±0.4	0.017±0.029	0.84±0.14	9.8±1.7	12.6±0.4	1±0.4	0.003±0.009	1.03±0.21	1.14±0.34
R 4	9.5±1.5	12.2±0.3	0.012±0.019	0.93±0.14	10.4±1.7	11.8±0.3	1.2±0.4	0.002±0.005	1.16±0.22	1.28±0.33
R 5	8±1.4	14.6±0.4	0.033±0.05	0.69±0.12	8.2±1.3	14.5±0.4	0.4±0.3	0.024±0.04	0.74±0.14	0.8±0.21
R 6	8.1±1.5	13.8±0.4	0.039±0.065	0.72±0.13	8.7±1.6	13.5±0.5	0.7±0.4	0.015±0.032	0.83±0.18	0.92±0.27
R 7	8.1±1.2	12.3±0.3	0.046±0.06	0.81±0.12	9.2±1.3	11.8±0.3	1±0.3	0.01±0.018	1.01±0.17	1.09±0.27
R 8	9±1.5	12.7±0.3	0.017±0.027	0.87±0.14	10±1.8	12.2±0.4	1.1±0.4	0.003±0.007	1.1±0.22	1.21±0.33
R 9	7.6±1.5	14.9±0.5	0.049±0.08	0.65±0.13	8.7±1.5	14.4±0.4	0.6±0.3	0.017±0.033	0.76±0.15	0.83±0.23
All	8.6±0.4	13.4±0.1	0.025±0.011	0.79±0.19	9.5±0.5	12.9±0.1	0.9±0.1	0.006±0.003	0.95±0.05	0.97±0.09

The similarities and differences between the ten replicates are striking. The hazard rate of each curve was estimated, and the MLE parameters of the Weibull and Gompertz distributions were estimated for the first 40% of deaths (Fig 3.1.b). All replicates showed an initial rapid increase in hazard rate followed by a marked deceleration. Each replicate appeared to be better fit by a Weibull than a Gompertz distribution (Table 3.2), except for replicate 5 for which the AIC criterion [Burnham and Anderson, 2004, Singer and Willett, 2003] appeared to marginally favor the Gompertz. The Weibull α, corresponding to the slope of the hazard log-log plot, was remarkably conserved between replicates (Table 3.2), with a mean of 8.5 and a standard deviation of 0.63, indicating that the environmental variation acting to differentiate replicates acted primarily to alter the timescale of death, leaving the coefficient of variation constant. The conservation of hazard curve shape between replicates was further confirmed in the quantile-residual plot (Fig. 3.2c), which demonstrates that the ratio between time-to-quantiles for each curve remain generally constant (forming roughly horizontal lines). Where the quantile-residual plot subtly deviated from horizontality, as in the case for Replicate 1, the quantile-residual line remained linear, suggesting a small variation in the Weibull α. For more information about Quantile-Residual plots, refer to 3.3.2)

The goodness of fit of the Weibull distribution and rough colinearity of the quantile-residual plot suggests that the curves differ primarily in their timescale, but that the shape of the curves was conserved between replicates. As such, we reasoned that the individual replicates might be well suited for aggregation using the AFT model proposed for device registration in chapter 2. The AFT model was solved using the log-linear regression model

$$\log(y_i) = \mu + b_i x_i + \varepsilon_i \tag{3.13}$$

which posits that (logs of) death times of replicates differ by an additive factor b_i. After obtaining the values for b_i for each replicate, the quantity $\varepsilon_i + \mu$ represents a "standardized" death time. The so-standardized data could then be aggregated into a single high-resolution aggregate curve containing 12,932 individuals (Figure 3.2e). On this data, the Weibull model fit to the first 40% of deaths was a vastly better fit than the Gompertz model according to the AIC criterion (Table 3.1). The quantile-residual plot of these MLE fits confirms this picture (Fig. 3.f).

3.5 Mortality Distributions and Extreme Value Distributions

Why would populations of *C. elegans* exhibit a Weibull distribution? One possibility might lie in Extreme Value (EV) theory, where Weibull distributions arise as the asymptotic limit of the distribution of maxima or minima of a set of independent and identically distributed (iid) random numbers [Fisher and Tippett, 1928]. A full treatment of the mathmatics behind EV theory is beyond the scope of this document–for a cogent overview refer to [Rinne, 2008]–so here we present a simple model and a numerical simulations to demonstrate the principles involved. Consider a random variable X representing the failure of a system component. Let X have a distribution function $F(x)$ and probability density $f(x)$. Consider further the minimum of a set of samples drawn from $f(x)$, i.e. a set of iid random variables X_i,
$\min(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$. This minimum, too, will be a random variable, call it X'_n. Examination of this quantity has many practical purposes. For example, consider a chain–the type that one might use to tie an angry bear to a tree. The chain is made of n links and, under load, each link fails at a time sampled from $f(x)$. The earliest of these times represents the failure of the first link, and, alas, of the entire chain. The distribution of $X'_n = \min(X_1, X_2, X_3, X_4, , X_n)$ therefore represents the distribution of failure times of the chain and determines, for example, the average time until the bear gets dinner (Fig. 3.3).

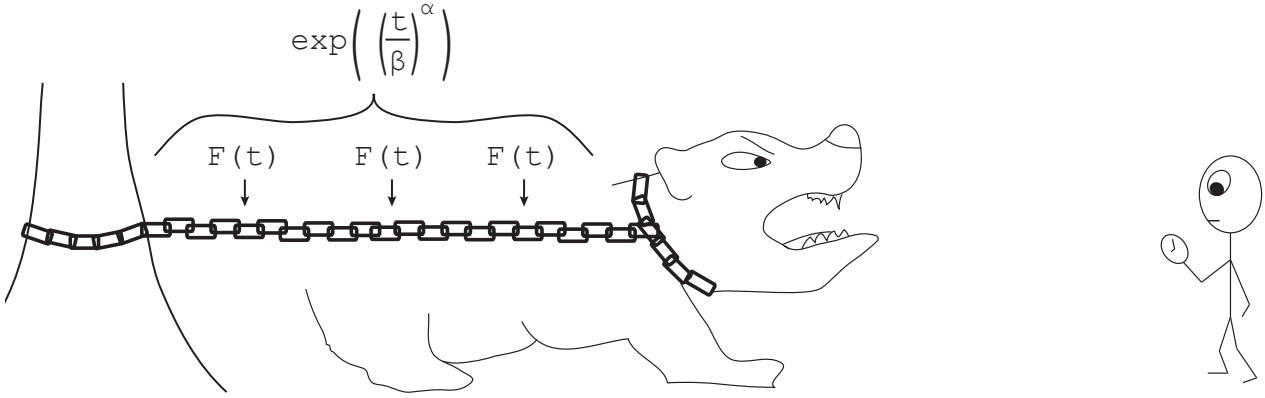

Figure 3.3: A physical model of an extreme value process A load-bearing (bear-loading) chain will fail when the first link breaks. A large class of distributions of link failure times will produce a chain whose failure time is distributed according to the Weibull distribution (shown).

One could imagine a variety of failure dynamics for individual links, for example

$$f(x) = \frac{1}{b-a} \quad \text{for } a \leq x \leq b, \text{ and } 0 \text{ otherwise} \qquad \text{(uniform distribution)} \qquad (3.14)$$

$$f(x) = b\exp(-bx) \qquad \text{(exponential distribution)} \qquad (3.15)$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad \text{(normal distribution)} \qquad (3.16)$$

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(\frac{(\log x-\mu)^2}{2\sigma^2}\right) \qquad \text{(log-normal distribution)} \qquad (3.17)$$

$$f(x) = \frac{a}{b}\left(\frac{t}{b}\right)^{a-1} \exp\left(-\left(\frac{t}{b}\right)^{a}\right) \qquad \text{(Weibull distribution)} \qquad (3.18)$$

These probability densities and their corresponding hazard functions are plotted in Fig. 3.4. The failure densities correspond to lifespan distributions below.

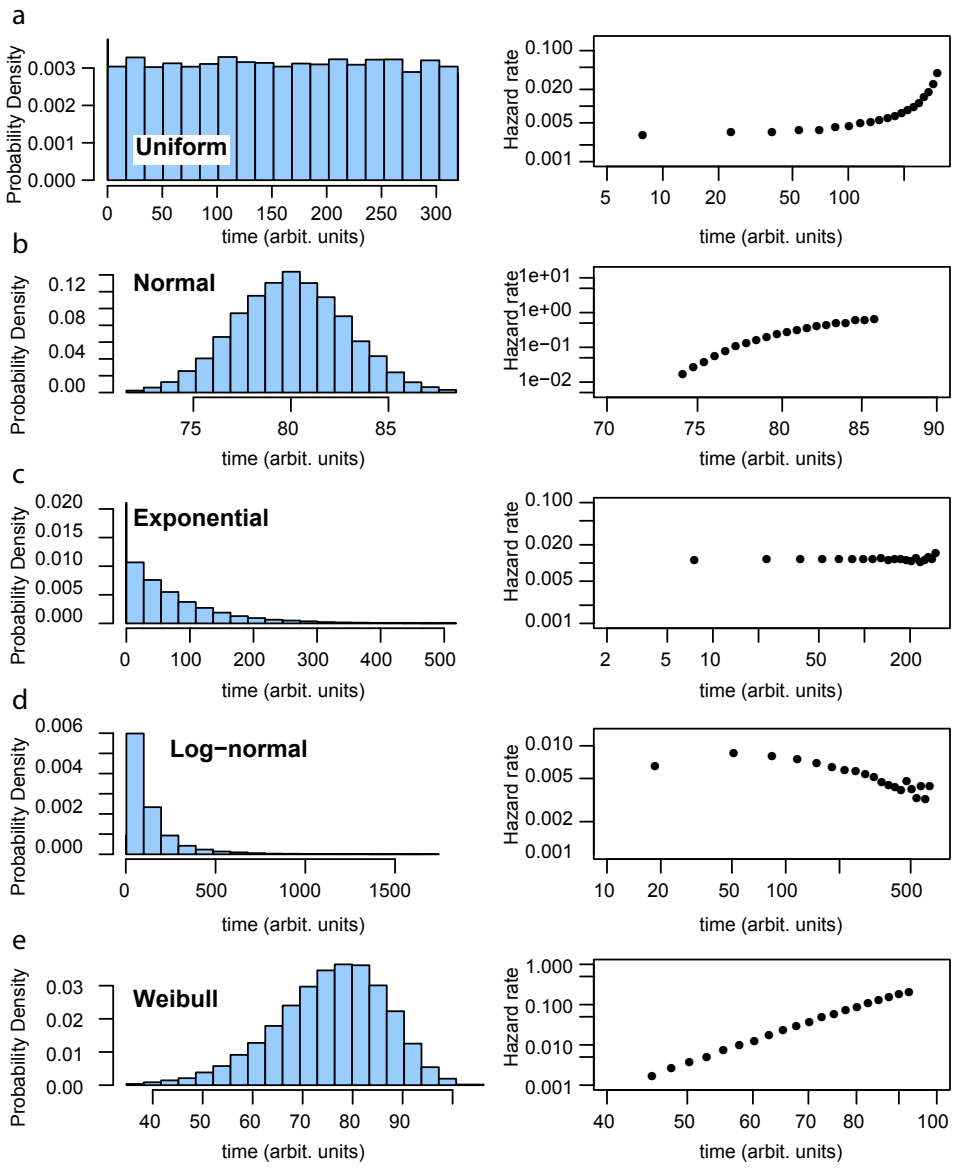

Figure 3.4: The shape of several types of failure distributions. Sets 10,000 random samples were drawn from a **(a)** uniform distribution **(b)** normal distribution (bounded between 0 and 300), **(c)** exponential distribution, **(d)** log-normal distribution, and **(e)** Weibull distribution. Each set is plotted as *(left)* a histogram and as *(right)* a hazard function estimate.

A simple simulation was run to demonstrate the behavior of minima of sets of so-distributed numbers. The following procedure was performed for each distributional form–N random samples were generating according to $f(x)$, and the minimum $\min(x_1, x_2, \ldots, x_N)$ was calculated. 10,000 such sets and their corresponding minima were obtained. The idea here is that each set of

N samples corresponds to a system of N parts, each failing according to $f(x)$. The minimum of the samples corresponds to the first component to fail. If the first failure "kills" the system, then the set of 10,000 minima calculated represents the lifespan distribution of a population of 10,000 such individuals. To this lifespan distribution corresponds a hazard rate, which was calculated for increasing N.

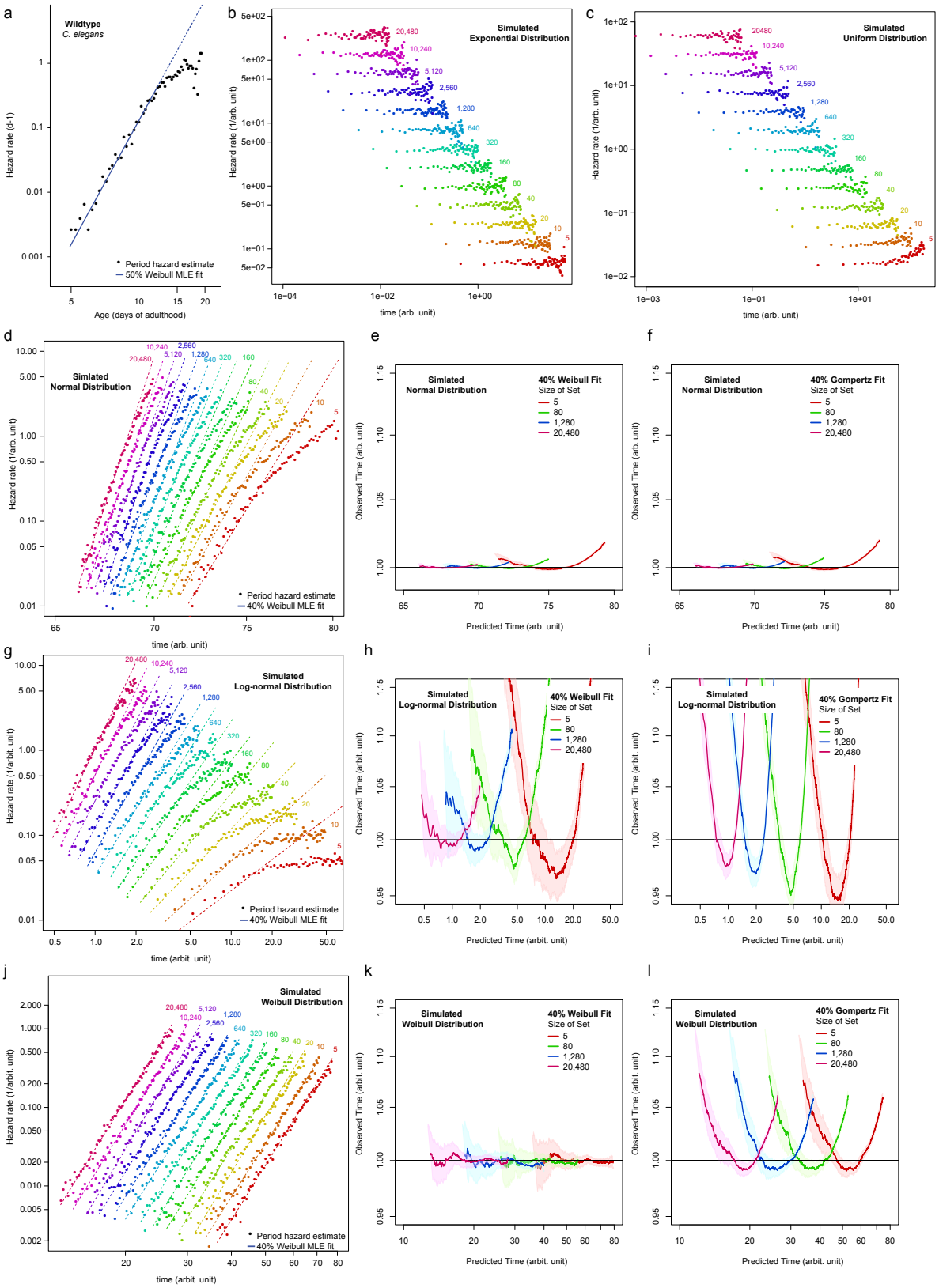

Figure 3.5: A simulation of Extreme Value phenomena. Caption is continued on the following page

Figure 3.5 (Cont'd): A simulation of Extreme Value phenomena. (a) The hazard rate for wild-type C. elegans, shown in Figure 2.1, is replotted for comparison to simulations of extreme value distributions. The minima were calculated for each of 10,000 sets of random samples drawn from a **(b)** exponential distribution, **(c)** uniform distribution, **(d-f)** normal distribution, **(g-i)** log-normal distribution, and **(j-l)** Weibull distribution. For each distributional form, the number of samples in each of the 10,000 sets was varied from 5 to 20,480 (colored red to purple, numbers shown). The hazard rate corresponding to each distributional form and population size was estimated and plotted, and MLE Weibull and Gompertz fits were calculated on the earliest 40% of deaths. The quality of these fits was demonstrated via quantile-residual plots comparing the MLE parametric fit against the simulated data for **(e-f)**, normal, **(h-i)** log-normal **(k-l)**, and Weibull distributed random samples. Quantile-residual plots were omitted for exponential and uniform distributed random samples, as the derived hazard rates are obviously poorly fit by either Weibull or Gompertz models.

This simulation demonstrates how lifespan distributions expose the signature of the failure statistics of components underlying lifespan. Exponentially and uniform distributions produced increasingly flat hazard functions as N increased (Fig. 3.5b-c). At the lower limit of a single component, $N = 1$, the failure dynamics of the entire system will match that of the single component–note how, for the uniform distribution, systems with $N = 5$ look very similar to Fig. 3.5.b, but converge to a constant hazard as N increases, as should be the case for an exponential failure distribution.

In contrast, normally-distributed and log-normally distributed component failure statistics produce hazard functions that look increasingly Weibullian at higher N (Fig. 3.5d-i). Weibull-distributed component failures produced Weibull hazard functions at all N, reflecting the well-known property that the Weibull distribution is closed under minima [Rinne, 2008].

Ultimately, the most surprising element about these models is how well they fit the early deaths actually observed in *C. elegans* lifespan distributions. The convergent behavior of systems involving minima provides a clear modality by which a variety of underlying failure statistics will tend to produce Weibullian failing systems. The underlying dynamics of worm death may be complicated. A nematode, for example, might die from vulval rupture, microbial invasion of the gut lining, hyper-proliferation of germ-cells, and so on. Each "disease" might be expected to demonstrate its own complicated time-dependent dynamics. To form an animal that dies

according to a Weibull distribution, these processes need not share the same mechanisms–a vulva is clearly different than the gut and will fail differently. To form an animal that dies according to a Weibull distribution, these processes need only have the same *failure distribution*, a phenomena we might expect evolutionary selection against weak components to bring about. This idea is discussed later in detail. First, we must take a more careful look at the tail of the distribution, the oldest living animals.

3.6 A simple model of heterogeneity accounts for the observed deceleration in hazard rate

Abundant evidence has been collected suggesting that individual *C. elegans* are quite heterogeneous, even within inbred populations housed in identical environments. Animals show initial [Pincus et al., 2011] and acquired differences in transcriptional markers [Zahn et al., 2007] and progression along stereotyped degenerative trajectories [Herndon et al., 2002, Huang et al., 2004]. Our analysis up until this point does not assume all individuals are identical. In fact it does quite the opposite—the hazard models explicitly describes the lifespan variation observed in populations. However, the models presented do assume that all individuals die according to a single underlying law, whose *parameters* are conserved across all individuals. We now consider a more general case, where we allow heterogeneity to exist between individuals in a form that causes their deaths to deviate from the single underlying law. The term "frailty" is used in a variety of contexts, but here we adopt the formal definition that describes the situation where the value of some parameter, for example the Weibull β, varies between subsets of individuals in a population. In Fig. 3.2b, we observed that independent replicates of wild-type *C. elegans* lifespan curves varied primarily according to their Weibull β parameter. Might individuals within a single experimental replicate vary in the same way? Attempts have been made to measure such variability directly [Pincus et al., 2011], and several candidate molecules exist that partially explain heterogeneous aging rates. However, these are hard to measure accurately in large populations, and, furthermore, many heterogeneous lifespan determinants may not be directly measurable at all using current experimental approaches. However, the heterogeneity might be identified phenomonologically through its effect on population-level statistics. To do this, we

chose perhaps the simplest model of population heterogeneity, a proportional frailty model [Aalen et al., 2008]:

$$h(t|Z) = Zh(t). \tag{3.19}$$

Here, Z represents the effect of some unmeasured heterogenity in a factor acting proportionally on the baseline hazard $h(t)$ We use "baseline hazard" to refer to the hazard that would be exhibited by populations homogenous in respect to Z. Since we observe that early deaths in wild-type *C. elegans* appear well-fit by a Weibull model, it seems reasonable to choose this as a baseline hazard. This gives us

$$h(t|\alpha, \beta, Z) = Z\frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1} \tag{3.20}$$

In this model, the parameters α and β are shared between all individuals, but Z varies between individuals. Z is by definition unmeasured, but we model it according to a probability distribution to understand its effects. The probability distribution must have a few properties–the most important of which is that it must always be greater than 0, as negative frailty values would produce negative hazard rates, which are nonsensical. A common choice is the gamma distribution, whose density is

$$G(x, k, \theta) = \frac{\theta^k}{\Gamma(k)}x^{(k-1)}\exp(\frac{-x}{\theta}) \tag{3.21}$$

This choice is not based on any fundamental theory on or observation of inter-individual diversity–it is chosen for mathematical convenience. The gamma distribution can represent a wide range of distributional shapes and has a clean Laplace transform, allowing it to produce a closed-form expression for the survival and hazard rates of heterogenous populations. Here, we set the mean of our frailty distribution to one–indicating that, on average, frailty has no effect on the hazard rate but merely contributes to its variance. Because a gamma distribution has a mean of k/θ, we fix this to 1 by setting $k = \theta$. This distribution then has a variance of θ.

The survival function based on the hazard model $h(t) = Zh(t)$ with Z distributed according to

$G(x, \theta, \theta)$ can be derived analytically [Aalen et al., 2008], taking the form

$$S(t, \theta) = \left(1 + \frac{1}{\theta} H(t)\right)^{-\frac{1}{\theta}}, \tag{3.22}$$

where $H(t)$ is the cumulative hazard up to t.

Using the Weibull model for a baseline hazard, $H(t) = (t/\beta)^{\alpha}$, we obtain the functions

$$S(t, \theta) = \left(1 + \frac{1}{\theta}\left(\frac{t}{\beta}\right)^{\alpha}\right)^{-\frac{1}{\theta}} \tag{3.23}$$

$$h(t, \theta) = \frac{\frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1}}{1 + \theta\left(\frac{t}{\beta}\right)^{\alpha}} \tag{3.24}$$

$$\log(h(t, \theta)) = \log(h(t)) - \log\left(1 + \theta\left(\frac{t}{\beta}\right)^{\alpha}\right) \tag{3.25}$$

We see that the addition of a gamma-distributed heterogeneity term produces a simple effect on the observed population hazard rate. The new hazard rate takes the form of the baseline Weibull hazard with a deceleration term $\log(1 + 1/\theta(t/\beta)^{\alpha})$ (Fig. 3.6a). The presence of heterogeneity in the hazard rate produces a mortality deceleration in the absence of any deceleration at the individual level. Were this phenomena at work in actual populations, hazard rate could be explained as a demographic phenomenon–the deviation of experimentally observed hazard from the underlying hazard experienced by individuals [Steinsaltz, 2005]. Furthermore, this deceleration phenomena does not depend on the exact form of the frailty model chosen. Frailty distributions including the inverse Gaussian, the positive stationary distribution, the power variance function, the compound Poisson distribution, and the log-normal distribution produce similar effects [Duchateau, 2008] Chapter 4. It is interesting to note that the log-logistic distribution, often used to model decelerating hazard rates, emerges as a special case of eq. 3.23 obtained when $1/\theta$ (the variance of the frailty distribution) is set to 1.

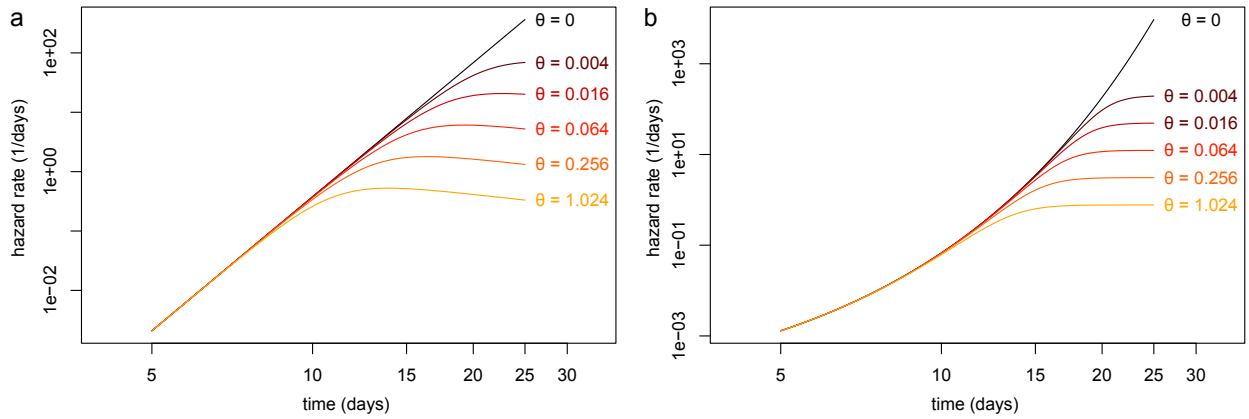

Figure 3.6: The effect of heterogeneity among individuals on a population's hazard rate. Hazard rate plots demonstrate the effect of incorporation of a gamma-distributed proportional frailty term into a **(a)** Weibull and **(b)** Gompertz model. The analytic expression for Weibull and Gompertz models are plotted for constant Weibull α and β and Gompertz a and b parameters while θ, the standard deviation of the gamma-distribution of frailties, is varied.

Other hazard models, for example, the Gompertz hazard, yield analogous results:

$$S(t,\theta) = \left(1 - \theta\frac{a}{b}\left(1 - \exp(bt)\right)\right)^{-\frac{1}{\theta}} \tag{3.26}$$

$$h(t,\theta) = \frac{a\exp(bt)}{\left(1 - \theta\frac{a}{b}\left(1 - \exp(bt)\right)\right)} \tag{3.27}$$

$$\log(h(t,\theta)) = \log(h(t)) - \log\left(1 - \theta\frac{a}{b}\left(1 - \exp(bt)\right)\right) \tag{3.28}$$

As before, the effect of adding a frailty term is to introduce a deceleration term to the baseline population Gompertz hazard, (Fig. 3.6b).

These relations demonstrate two important points. First, heterogeneity of the sort that produces a proportional frailty term is expected to produce an apparent deceleration in a populations hazard without resorting to specific biological processes. Second, for reasonable distributions of frailty with small variance, early deaths should closely approximate the underlying individual mortality law (Weibull, Gompertz, etc.).

The closed form solutions to frailty models based on Weibull and Gompertz distributed deaths allows us to obtain MLE estimates of the three-parameter distributions (Table 3.2). The MLE

models including frailty yielded better fits to the data than their non-frailty counterparts (Fig. 3.2e-g), allowing the full range of data (100% of deaths) to be well-approximated in each wild-type experiment and in the aggregate. The AIC values resulting from the MLE fits of individual replicates suggested that the Weibull frailty model fit better than the corresponding Gompertz frailty model, with the exception of Replicate 9 which was marginally better fit by the Gompertz frailty model (Table 3.1). In the 13,000-individual aggregate set, comparison between AIC values suggested a better fit with the Weibull frailty model than Gompertz. Note how the observed hazard data appears linear on a log-log plot. The Gompertz model deviates qualitatively from this, bending upward. The addition of a frailty term allows the Gompertz model to ”bend” back to linearity, therefore essentially mimicking the Weibullian linear behavior, but it does not do so in a way that allows it to quantitatively outperform the Weibullian frailty model.

Given the previous result showing that the early 40% of deaths are better fit by Weibull distributions, it is understandable that Weibull frailty models performed better than Gompertz frailty models. The two models are most easily distinguished among the earliest-dying cohorts, which are also the least affected by the frailty term (Fig. 3.6). The better fit can be seen both with and without frailty, as the Gompertz and Gompertz Frailty models deviate more from horizontality in the quantile-residual plots, among deaths occurring between 5 and 10 days. The quantile-residual plot in Fig 3.1. shows that even though the addition of a Gamma-distributed frailty term drastically improves the fit of the models, it does not perfectly fit. The frailty models appear to predict a faster death rate than observed in the age range between 15 and 18 days, perhaps suggesting that the actual frailty distribution is not gamma-distributed, or that proportionality assumptions required for our AFT-regression are not fully met at the tail of these distributions.

It is important to note that the existence of deceleration phenomena introduces fundamental difficulties in differentiating between Weibull and Gompertz models of the hazard rate. The two distributions are most easily distinguished by comparison of their tails–during early and late deaths. In our experimental data, differences in late deaths are masked by deceleration phenomena. Differences in early deaths may also be difficult to evaluate precisely, as these occur during the interval when non-aging related deaths are known to occur–vulval ruptures and ”bag

of dead egg" phenotypes are usually observed early in adulthood. Such "non-ageing-related" deaths might mask the underlying ageing dynamics at early time points. Such phenomena can be reduced by setting appropriate experimental conditions (at 25 °C, vulval ruptures occur at negligible rates), and their statistical consequences mitigated by appropriate censoring(as is done in our data). However, the current experimental reality means that differentiation between Weibull and Gompertz models based on quantitative results such as AIC should be taken with a grain of salt; the relative performance of the two models may depend strongly on censoring-related issues, or mechanisms biologically distinct from aging at work at the tails of the distribution. The differentiation between models based on our data set is ultimately a qualitative one. Our *elegans* data suggest hazard rises according to the logarithm of time, in agreement with the Weibull model. The Gompertz model can be made to approximate the Weibullian behavior through introduction of a frailty term, but requires introduction of greater underlying heterogeneity (a larger θ) than would a Weibull model in these circumstances.

3.7 Discussion

Wild type *C. elegans* death times appear well-fit by Weibull model across the first 40% of deaths, and well fit across all deaths by a Weibull model including a frailty term (Fig. 3.2. The observed dynamics are therefore consistent with a simple model by which *C. elegans* die according to some Extreme Value (EV) phenomena (Fig. 3.5). One requirement for an EV process to produce a Weibull distribution is that each component necessary for life exhibits a similar distribution of failure times. One might imagine evolution producing this type of failure architecture, as any component that fails so early as to effect an animal's fitness will be selected against. In this way, any "weakest-link" that appears through random variation will be selected against, resulting in a system where the failure distributions of important components are balanced. The data presented in this chapter are merely consistent with this model rather than definitively proving it, however the surprising simplicity of the mathmatical laws observed in the experimental data suggest a single unifying principle at work. These issues are discussed in more detail in the discussion chapter.

Chapter 4

The temperature scaling of *C. elegans* lifespan between 20 °C to 36.5 °C

4.1 Temperature and Lifespan

Temperature is a major determinant of lifespan in most ectotherms, including *C. elegans* [Klass, 1977], *Drosophila melanogaster* [Miquel et al., 1976]. Succinctly, individuals live shorter at high temperatures. This trend appears to hold even among mammals, where experiments perturbations to temperature regulatory mechanisms of mice [Conti et al., 2006] suggested that lowering body temperature might extend lifespan as a general phenomenon in living systems. This temperature dependence of lifespan is widely ascribed to the temperature dependence of the reaction rates of various biochemical reactions [Somero and Hochachka, 1968, Isied et al., 1984, Sekiya et al., 2009, Chaui-Berlinck et al., 2004]. In 1928, Raymond Pearl used his investigations into the lifespan of fruit fly and cantaloupe seedlings to suggest that "life varies inversely as the rate of energy expenditure during its continuance" [Pearl, 1928]. Essentially, animals start out with some limited quantity, which he calls vitality, and the faster an animal performs, the sooner it runs out and dies. Temperature might shorten lifespan in exactly this way–speeding up animals and consequently shortening their lives.

What sort of reactions might be accelerated by temperature? Since metabolic reactions run faster at higher temperatures, they are one likely candidate. Certainly, the rate of metabolism correlates

with lifespan between species [Ruggiero and Ferrucci, 2006] though with many exceptions [Miller and Austad, 1999]. A prominent hypothesis assigns cumulative molecular damage as the mechanism linking the rate of metabolism and lifespan–radical oxygen species created as unavoidable by-products of respiration were hypothesized to cause irreparable damage to DNA and proteins [Walker et al., 2005]. Perturbations that decreased the rate of ROS creation (lowered metabolism rate by temperature or starvation, activation of ROS scavenging or repair mechanisms) would delay the time at which cumulative exposure caused death, whereas perturbations that increased ROS creation (increased metabolism, mitochondrial electron transport chain decouples such as paraquat, mutations in repair mechanisms) would accelerate death. Definitive evidence of this mechanistic hypothesis failed to materialize in the mouse [Zhang et al., 2009],nematode [Raamsdonk and Hekimi, 2012] sod4 multiple knockouts), however, and other mechanisms have been proposed, including effectors of protein homeostasis [Ben-Zvi A, 2009, David et al., 2010]. Regardless, all such proposed mechanisms share the common idea that the temperature dependence of lifespan results from the temperature dependence of lifespan-determining biochemical reactions.

In *C. elegans*, lifespan is studied at two temperature extremes. Housed at temperatures between 15 °C and 25 °C, animals die over the course of several weeks, with lifespan determinants widely considered to affect “aging”. Housed at temperatures between 34 °C and 36 ° *elegans* live for less than twelve hours and lifespan determinants are generally considered to effect “stress resistance”. However, these two temperatures delimit a continuous spectrum and the question arises whether aging and stress resistance do in actuality represent qualitatively distinct phenomena. Some evidence does exist to differentiate between the two regimes. For example, a subset of genetic determinants of lifespan differs between the two temperature regimes, for example *hsf-1* is required for normal lifespan at 25 °C but not 35 °C [McColl et al., 2010]. However, many genetic determinants of lifespan are also shared between the two temperature regimes, in particular insulin signaling pathway components whose disruption can increase survival times at both low and high temperatures [Lithgow et al., 1995]. This shared genetics is often expressed by stating that lifespan and stress resistance appear closely coupled. However, death at these two temperatures may have more in common than their genetic determinants. We take as a cue our

initial experiments using the lifespan machine to study the hazard rate of animals dying at 35 °C (Fig. 2.5). The striking similarity of the hazard rate shape at both 25 °C and 35 °C is highly suggestive of some shared stochastic mechanisms underlying lifespan at both temperatures.

In chapter 3, we discussed briefly the relationship between the failure dynamics of multi-component systems and the failure of individual components, and showed how Weibullian lifespan distributions for systems might arise. Varying temperature would also appear a potentially fruitful means for testing such theories. One might imagine that increases in temperature might alter the underlying failure dynamics of components. Temperature shifts provide an experimental means for altering the rates of chemical reactions that might drive the failure of important components, or even introduce novel failure mechanisms. Temperature-based interventions may not be as targeted as, for example, genetic perturbations, but represent a good starting point. Given the intrinsic temperature dependence of many genetic perturbations themselves (*glp-1* [Arantes-Oliveira et al., 2002] *hsf-1* [McColl et al., 2010]), understanding such effects in wild type seems a natural first step.

4.2 The statistical form of lifespan distributions is the same at 25 °C and 33 °C

The lifespan machine allows acquisition of very-high resolution mortality at high temperatures (chapter 2). To evaluate the survival curve shape at high temperatures, we acquired 4,700 age-sychronous wild-type animals by hypochloride treatment and placed them on NEC937 bacteria (*E. coli* OP50 ΔuvrA::KanR) at 20 °C. Animals were maintained at 20 °C, and transferred as late-l4 larvae onto fresh plates containing 10 mug/mL FuDR. To separate the effects of temperature on aging from its effects on development, animals were retained at 20 °C unil the second day of adulthood, at which time they were were transferred onto flatbed scanners calibrated to operate at 33 °C. Their death times were obtained through automated analysis of captured image data, using the techniques described in chapter 2.

Analysis of this data confirmed the initial results shown in Fig 2.5a-b. Six hours after transfer to 33 °C, the hazard rate became measurable and exhibited a rapid increase after ten hours of

exposure (Fig. 4.1. Maximum Likelihood Estimates of the Weibull and Gompertz parameters to the first 40% of deaths showed the Weibull to be the better model (ΔAIC = 104.7). We found lifespan at 33 °C to be well-fit by Weilbull and Gompertz models that include gamma-frailty terms, similar to our finding at 25 °C. Strikingly, the estimated Weibull α (Table 4.2), corresponding to the coefficient of variation in the population, fell within the range of values seen among deaths observed at 25 °C (Fig. 3.2b and Table 3.2). There are many reasons why the coefficient of variation might be the same at 25 °C and 33 °C, the most provocative of which is that the molecular mechanisms underlying diversity in lifespan at both temperatures are shared.

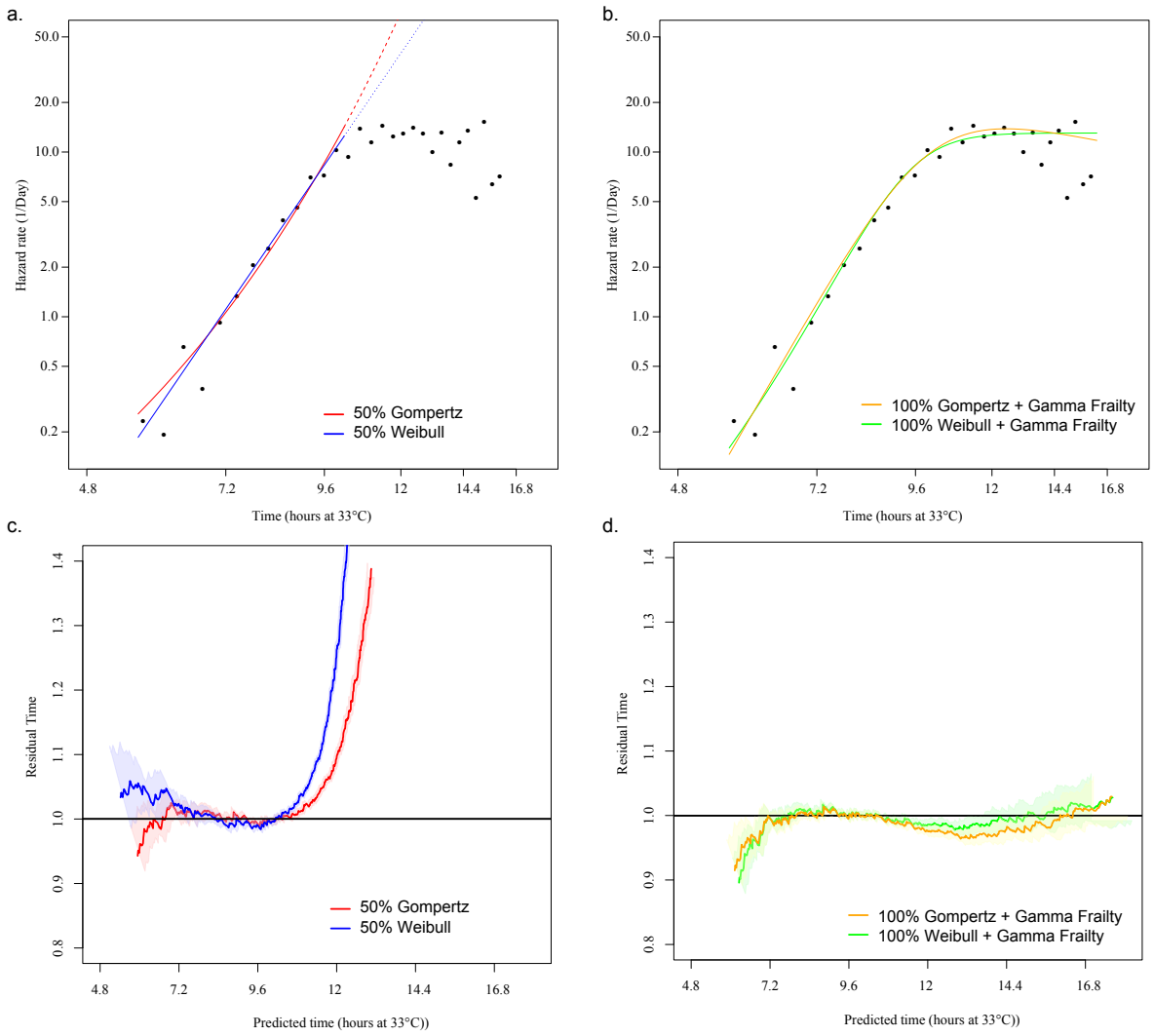

Figure 4.1: Wild-type mortality at 33°C. 4754 wild-type C. elegans nematodes were transferred to 33°C on their second day of adulthood, and their survival monitored by the lifespan machine. Death times were used to estimate the population hazard rate, which is shown with **(a)** MLE Weibull and Gompertz models fit to the first 50% of deaths and **(b)** MLE Weibull and Gompertz models incorporating a gamma-distributed frailty term fit to 100% of deaths. The quality of each MLE fit is demonstrated in the quantile-residual plots pictured both **(c)** without and **(d)** with frailty terms. For a description of quantile-residual plots, refer to 3.3.2. Experiments were performed by ZN and WA

Table 4.1: Summary statistics describing wild-type lifespan at 33°C. Deaths were collected across ten scanners in a single replicate, in which wild-type animals were transferred from 20°C to 33°C on the second day of adulthood. The relative performance of Weibull vs. Gompertz models was not consistently determined across individual scanners for the first 50% of deaths, however in the larger-population-size aggregate, Weibull did substantially better. The addition of a Gamma-distributed frailty term increased the performance of the Weibull relative to the Gompertz on the larger-population-size aggregate.

Device	Deaths	Censored	Median Lifespan	Delta AIC Gompertz v. Weibull	Delta AIC Gompertz Frailty v. Weibull Frailty
worm	322	126	0.40	-12.43	-0.61
teal	486	47	0.47	7.24	32.41
pink	499	51	0.42	0.61	32.99
maple	476	32	0.40	-3.80	-3.78
green	531	19	0.47	-11.99	-6.28
cyan	481	52	0.46	-3.83	4.23
blue	500	59	0.37	3.05	59.12
birch	484	22	0.40	-2.78	55.08
bear	484	30	0.41	10.11	1.09
azure	491	30	0.48	-6.31	-8.29
All Device Aggregate	**4754**	**468**	**0.43**	**15.58**	**178.53**

Table 4.2: MLE Parameters @ 33 °C. The maximum-likelihood estimate model parameters are provided for each model fitted, with 95% confidence intervals shown.

	Weibull Shape	Weibull Scale	Gompertz Rate (x103ˆ)	Gompertz Shape	Weibull Frailty Shape	Weibull Frailty Scale	Weibull Frailty Theta	Gompertz Frailty Rate (x103ˆ)	Gompertz Frailty Shape	Gompertz Frailty Theta
worm	8.4±2	0.4±0	1.11±2.772	24.68±5.88	9.3±2.5	0.4±0	0.9±0.6	0.304±1.155	30.01±8.68	34.35±0.39
teal	6.5±1.4	0.5±0	5.903±9.36	16.11±3.59	8.2±1.9	0.5±0	1.2±0.6	0.564±1.773	24.14±7.01	27.64±0.52
pink	8.7±1.9	0.4±0	0.843±1.91	23.61±5.2	10.9±2.9	0.4±0	1.8±0.8	0.036±0.2	34.79±10.54	40.06±0.66
maple	9.4±2.1	0.4±0	0.47±1.208	26.77±6	10.4±2.8	0.4±0	1.1±0.7	0.086±0.39	33.17±9.57	37.95±0.43
green	8.1±1.7	0.5±0	1.237±2.449	20.13±4.22	9.1±2.1	0.5±0	1±0.5	0.216±0.744	25.83±7.01	29.33±0.4
cyan	8.1±1.8	0.5±0	1.257±2.675	20.55±4.59	9.2±2.4	0.5±0	1.1±0.6	0.175±0.715	27.09±8.07	31.13±0.48
blue	10.2±2.1	0.4±0	0.299±0.767	30.36±6.35	12.7±3.5	0.4±0	1.7±0.9	0.003±0.033	47.07±15.18	54.66±0.7
birch	8.2±1.8	0.4±0	1.379±2.933	23.53±5.2	9.8±2.4	0.4±0	1.2±0.6	0.104±0.45	33.19±9.53	37.96±0.51
bear	8.3±1.8	0.4±0	1.518±3.093	22.81±4.93	9.2±2.7	0.4±0	1±0.7	0.163±0.88	30.89±10.85	36.31±0.54
azure	6.6±1.5	0.5±0	4.164±7.219	16.35±3.76	7.1±1.7	0.5±0	0.7±0.5	1.677±4.045	19.57±5.37	22.25±0.34
All	7.7±0.6	0.4±0	1.981±1.151	20.95±1.54	9.4±0.8	0.4±0	1.2±0.2	0.166±0.157	29.63±2.72	30.99±0.16

We then asked whether the similarities in survival curve shape observed between 25 °C and 33 °C existed at intermediate temperatures. We built a large imaging cluster consisting of 50 flatbed scanners to measure the lifespan of animals at small fraction of degree intervals across the temperature range of 20 °C to 36.5 °C. Over the course of five months, 1,750 age-synchronous larvae were obtained each Monday (as described previously) and on the second day of adulthood (Friday) they were loaded across ten scanners located in a single incubator. Scanners were calibrated using high-precision thermocouples, with scanner temperatures re-measured at the conclusion of each assay to confirm temperature stability. Because individual scanners demonstrated a natural spread in temperature around the central set point, often ranging +=1 °C, we were able to test several closely-spaced temperatures in a single experiment. According to this procedure, we explored the temperature region between 20 °C to 36.5 °C over the course of 25 experiments, run in the following chronological order: 34.5,30, 20, 28, 31, 22.5, 25, 36.5, 29.5, 32.5, 33.75, 28.25, 31.75, 28.75, 34, 28.9, 35.5, 32.2, 26, 27, 31.45, 29.5, 30.75, 35.1, and 25 °C. One experiment, the first 25 °C replicate, produced results inconsistent with other experiments at this temperature (Fig. 3.2) and was discarded.

4.3 The statistical form of lifespan distributions appears invariant at all temperatures 20 °C and 36.5 °C

We found that the shape of the survival curve remained remarkably stable across the entire temperature range as evidenced by a constant ratio of time-to-quantiles at each temperature (Fig. 4.2a). Departures from horizontality of the quantile-residual plots generally took the form of a non-zero slope but nonetheless linear ratio of time-to-quantiles, suggesting small variations in the Weibull α between replicates. The variation in these αs is consistent with that observed between independent replicates performed at the same temperature (25 °C, Fig. 3.2b). The hazard rate estimates are shown for the entire range (fig 4.3), with 50% Weibull MLE fits over-layed (Fig. 4.5), and as close-ups on representative curves from various temperature intervals (Fig. 4.5).

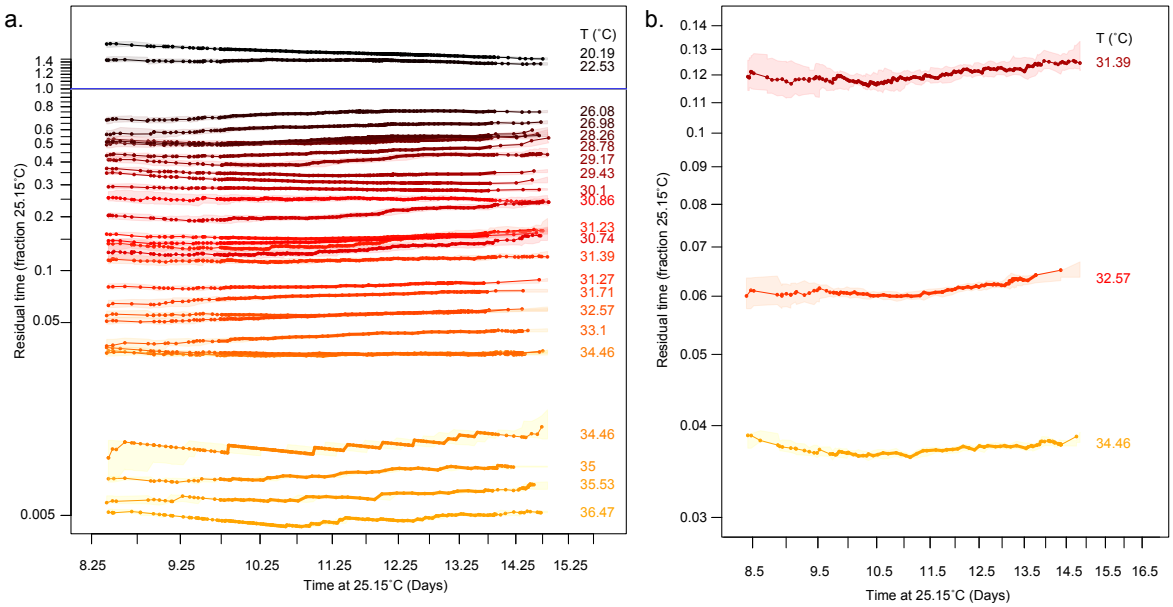

Figure 4.2: The quantile-residual plot for populations dying at temperatures ranging between 20°C and 26.5°C. A survival curve collected at 25°C was chosen as the reference against which the **(a)** quantile-residual plot for survival at all other temperatures were plotted. In **(b)** three representative curves were selected and plotted to demonstrate the apparent constant hazard ratio between lifespan across the entire temperature range. For a description of quantile-residual plots, refer to 3.3.2. Experiments were performed by ZN, ILM, WA, CW, and NES

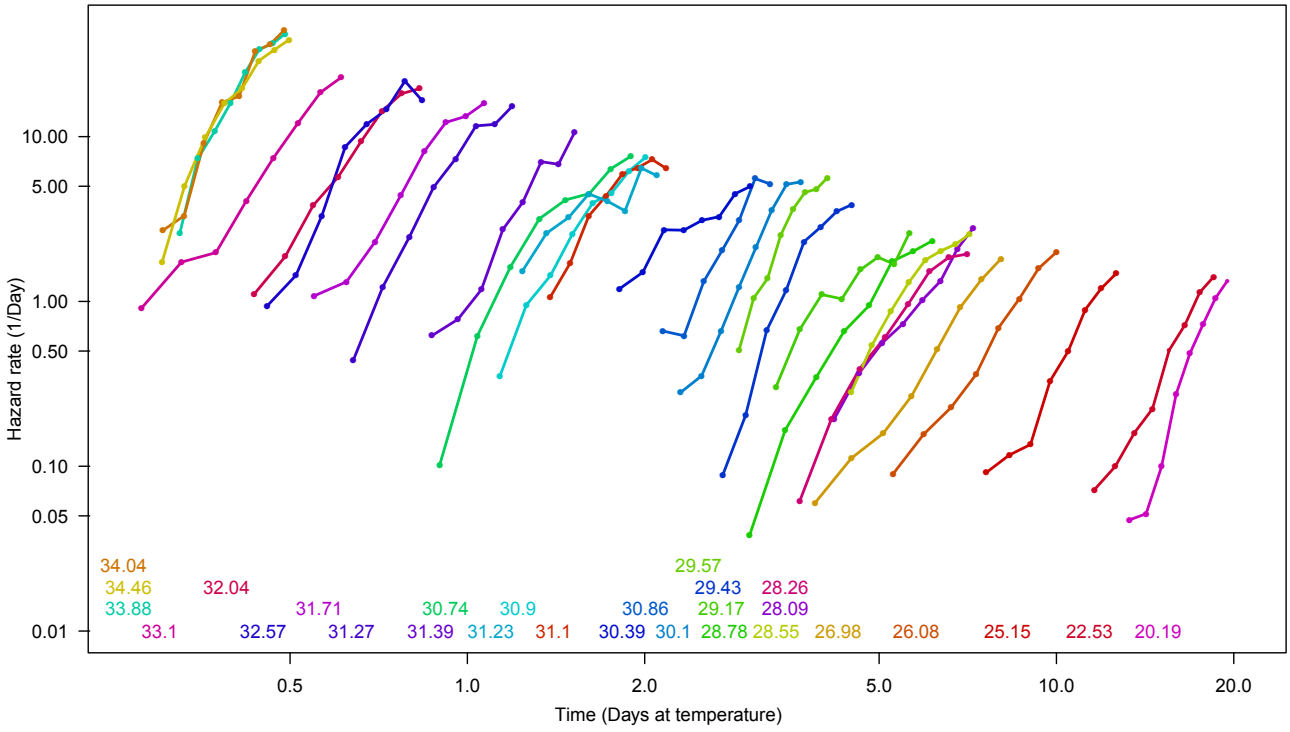

Figure 4.3: Temperature scaling of the wild type *C elegans* hazard rate above 20°C (piecewise-exponential estimate). Wild-type *C. elegans* nematodes were transferred on the second day of adulthood to a range of temperatures ranging from 20°C to 36.5°C, and their survival at that temperature monitored using the lifespan machine. Each curve represents the aggregate of several scanners each run in a single experimental replicate, using the AFT-regression model described in Chapter 2. Aggregation was performed to produce populations large enough to generate accurate hazard rate estimates. Experiments were performed by ZN, ILM, WA, CW, and NES.

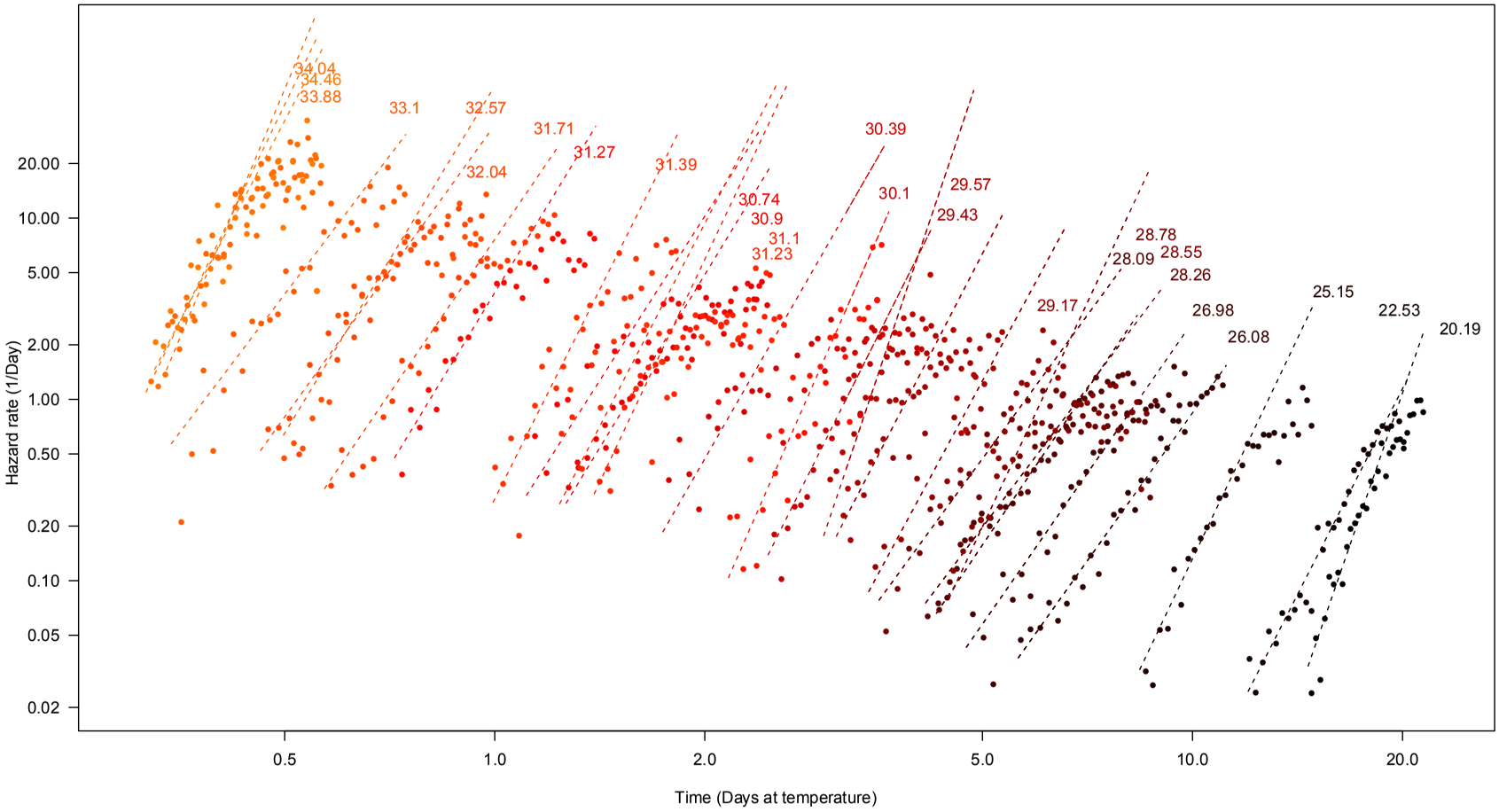

Figure 4.4: Temperature scaling of the wild type *C elegans* hazard rate above 20°C (MLE Weibull estimate). Caption is continued on the following page

Figure 4.4 (Cont'd): Temperature scaling of the wild type *C elegans* hazard rate above 20°C (MLE Weibull estimate). Wild-type *C. elegans* nematodes were transferred on the second day of adulthood to a variety of temperatures ranging from 20°C to 36.5°C, and their survival at that temperature monitored using the lifespan machine. Each curve represents the aggregate of several scanners each run in a single experimental replicate, using the AFT-regression model described in Chapter 2. Aggregation was performed to produce populations large enough to generate accurate hazard rate estimates. ML-estimate Weibull fits were obtained for the first 50% of deaths, shown as lines. Experiments were performed by ZN, ILM, WA, CW, and NES

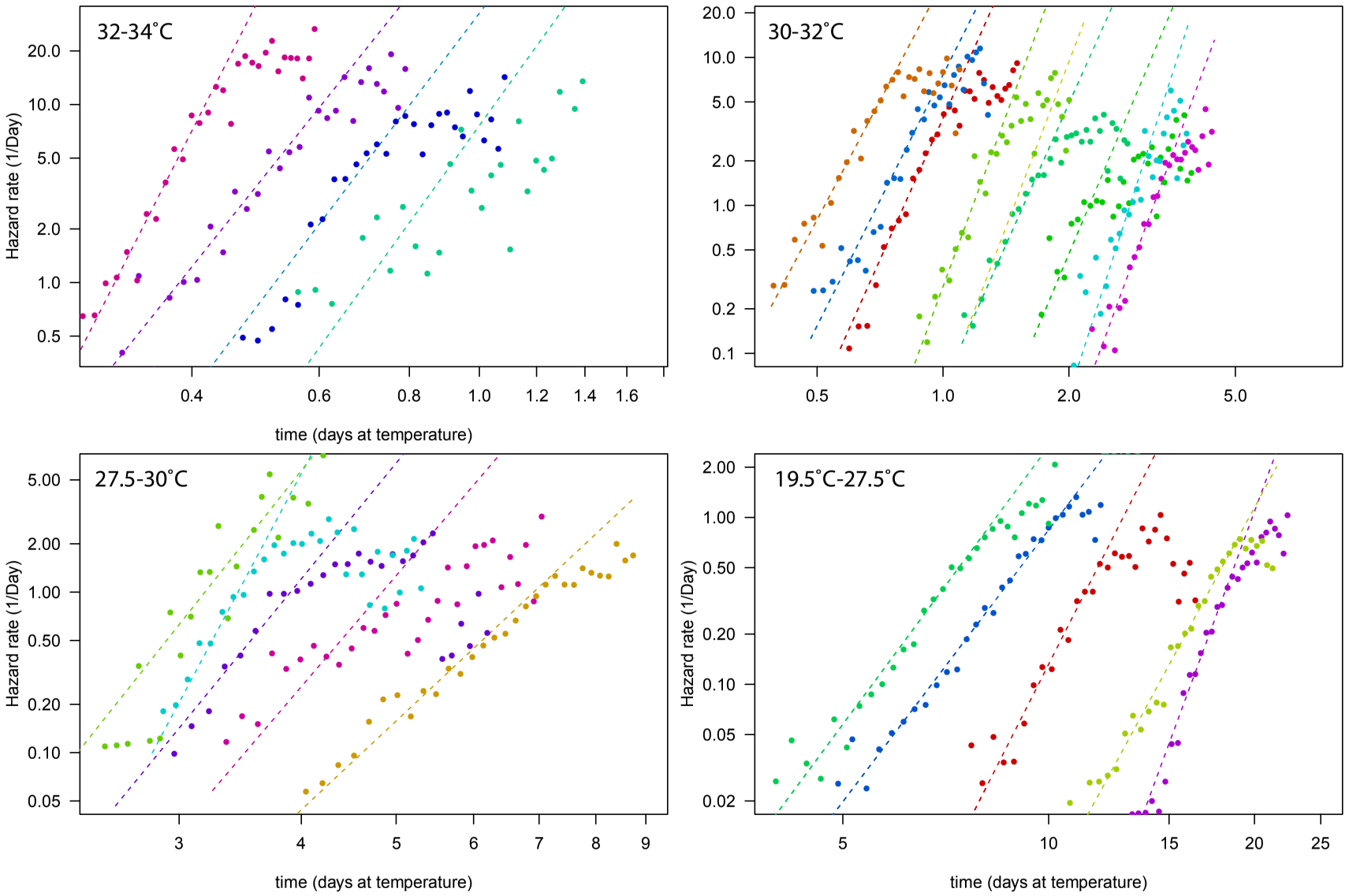

Figure 4.5: Temperature scaling of the wild type *C elegans* hazard rate above 20°C (Highlight of selected curves).
Caption is continued on the following page

Figure 4.5 (Cont'd): Temperature scaling of the wild type *C elegans* hazard rate above 20°C (Highlight of selected curves). Wild-type *C. elegans* nematodes were transferred on the second day of adulthood to a variety of temperatures ranging from 20°C to 36.5°C, and their survival at that temperature monitored using the lifespan machine. A small number of populations were selected across the temperature range, to allow close inspection of this subset of the data. Each curve represents the aggregate of several scanners each run in a single experimental replicate, using the AFT-regression model described in Chapter 2. Aggregation was performed to produce populations large enough to generate accurate hazard rate estimates. ML-estimate Weibull fits were obtained for the first 50% of deaths, shown as lines. Experiments were performed by ZN, ILM, WA, CW, and NES

Across the range of 20 °C to 36.5°C, the worm's average lifespan drops more than one hundred fold, from eighteen days to less than two hours. However, rather than decreasing smoothly across this temperature range, we found that lifespan appeared to follow four distinct scaling regimes, each specifying a distinct relationship between lifespan and temperature 25 °C–30 °C, 30 °C–31.5 °C, 35 °C–34.5 °C, and 35 °C and above (Fig. 4.6). Temperatures below 25 °C appeared to follow a fifth scaling relationship, but not enough data was collected at these temperatures to permit detailed study.

Strikingly, each individual regime appeared well modelled by the relationship $-\log(\beta_T) = c_1 + c_2\frac{1}{T}$, where β_T is the Weibull scale parameter measured for animals dying at temperature T, and c_1 and c_2 are constants. This relationship (discussed in 4.4) had been noted for temperatures between 15 °C–20 °C [Suda et al., 2012], but the segmented, piecewise-exponential relationship observed above 25 °C came as a complete surprise.

4.4 The Arrhenius Relationship

Investigations into the quantitative effects of temperature changes to chemical reactions have a long history of both experimental and theoretical approaches [Laidler, 1984]. The two met in the Arrhenius equation.

$$k = A\exp\left(\frac{-E_a}{RT}\right) \tag{4.1}$$

This equation describes a relationship between the rate constant of a chemical reaction k at temperature T, in terms of two free parameters A and E_a, and the Boltzmann constant R. The

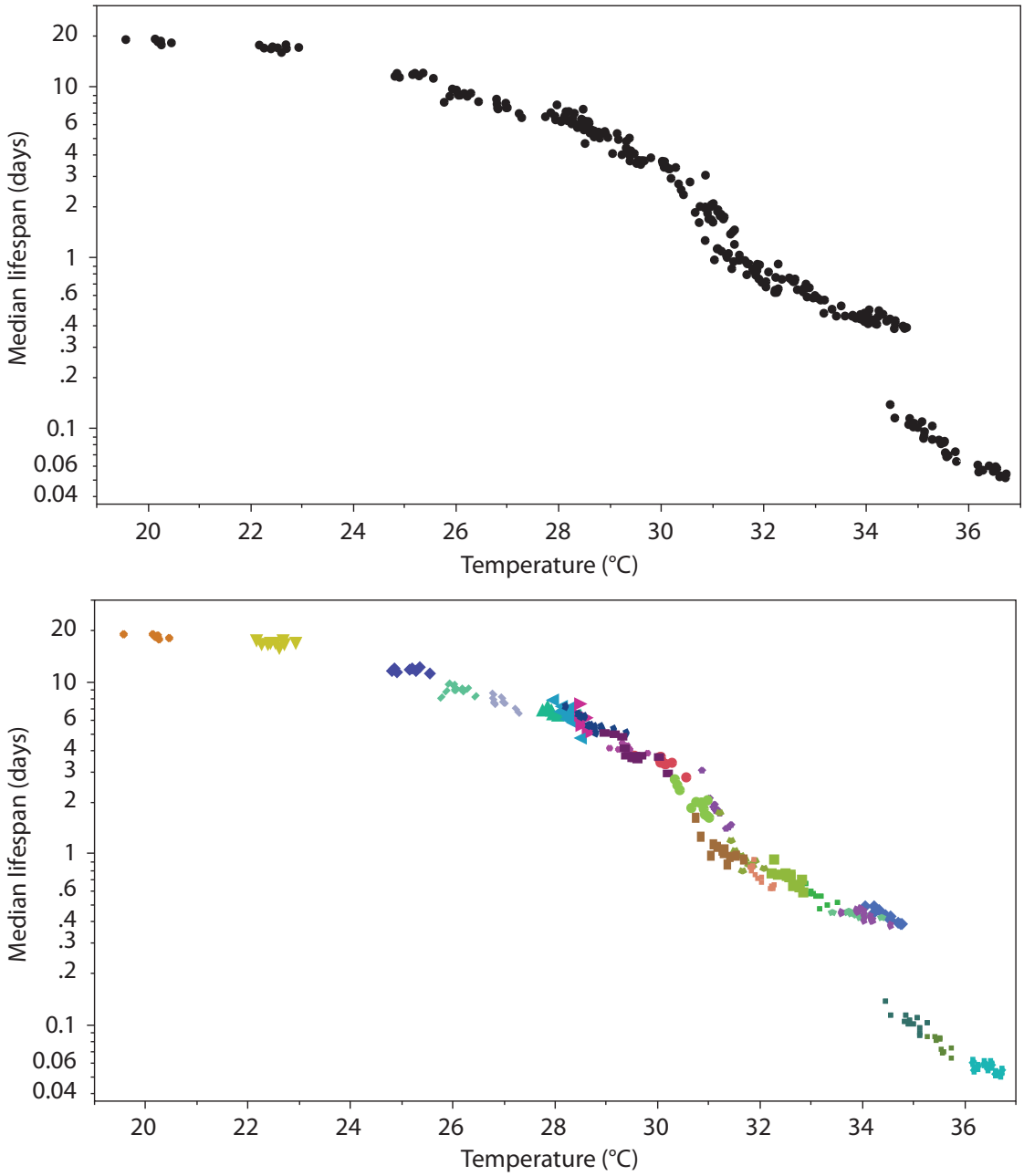

Figure 4.6: The scaling of median lifespan between temperatures 20 °C–36.5 °C. A survival curve was obtained for animals observed on each scanner in each replicate at each temperature, each curve containing between 100 and 250 individuals. The median lifespan was calculated for each survival curve and presented *(top)* without respect to the experimental replicate and *(bottom)* colored by experimental replicate. This second plot demonstrates the level of constancy between experimental replicates at the same or neighbouring temperatures. Experiments were performed by ZN, ILM, WA, CW, and NES

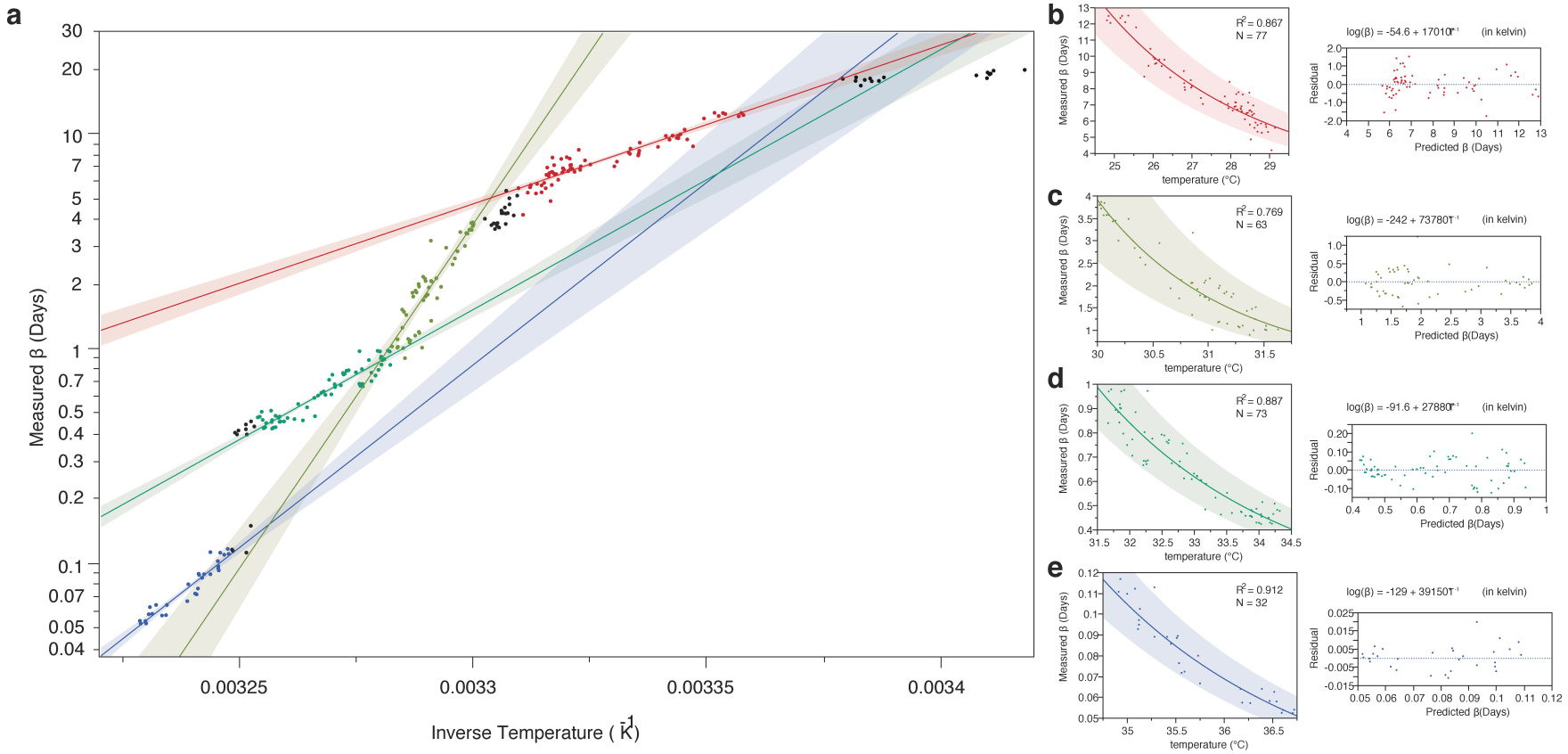

Figure 4.7: An Arrhenius plot for wild-type *C. elegans* lifespan. Caption is continued on the following page

Figure 4.7 (Cont'd): An Arrhenius plot for wild-type *C. elegans* lifespan. MLE Weibull parameters were calculated for mortality distributions collected across the range of 20°C to 36.5°C. **(a)** The natural logarithm of the Weibull β parameter was plotted against the inverse of temperature, revealing four distinct scaling regimes, shown individually in insets **(b-e)**. The residual of linear least-squared fits of the transformed variables were fit to each region, shown as straight lines, and the residuals were plotted *(right)* to demonstrate the quality of fit. Experiments were performed by ZN, ILM, WA, CW, and NES

equation was proposed because of its good fit to empirical data, but was later derived from first principles. Useful textbooks have been continuously written on this subject for over a century [Sethna, 2006, Atkins, 2001]. Yet, to briefly summarize, the distribution of energies of particles within an ideal gas can be derived from basic statistical principles. These results can be extended to molecules interacting in solution. These reactions occur only when sufficient energy is available in a collision between molecules to overcome the activation energy E_a. The fraction of particles in a group possessing the necessary energy can be shown to produce the relationship above. Increasing the temperature of a chemical system increases the fraction of particles possessing an energy greater than the activation energy E_a, thereby making reactions more likely.

The Arrhenius relationship has been modified in several ways, for example introducing a temperature dependent pre-exponential factor A. However, within narrow temperature ranges, the the Arrhenius relationship represents a good approximation to many reactions, and provides a means to estimate the activation energy of a reaction using on experimental data.

$$\log k = \log A - \frac{E_a}{R}\frac{1}{T} \tag{4.2}$$

Plots of $\log k$ against $1/T$ will be linear and the slope of the line will be E_a/R.

4.5 A simple model describing a rate-limiting reaction that determines lifespan

C. elegans nematodes naturally require a multitude of biochemical reactions in order to remain alive. However, it is not impossible that a single reaction, or a reaction pathway acting as a unit,

might act as a rate-limiting step in processes that affect aging. This hypothetical reaction would in effect determine the rate of living of an animal. To illustrate how such a reaction might act within the Arrhenius framework, we construct the following model

$$P \xrightarrow{k} C_{\text{poison}}. \tag{4.3}$$

The exact molecular mechanism of this reaction is unknown, but the reaction might describe for example an electron transport chain component `P` that reacts with a oxygen to produce a superoxide radical C_{poison}. Or, the reaction could describe a protein `P` that aggregates to produce an insoluble product C_{poison}. While the reaction is out of equilibrium, with the concentration of C_{poison} negligible relative to that of `P`, $\texttt{C}_{\texttt{poison}}$ will accumulate over time. We define some concentration of C^*_{poison} that would kill the animal–i.e the animal can cope as long as $\texttt{C}_{\texttt{poison}} \leq \texttt{C}^*_{\texttt{poison}}$, but above that level it dies. The time at which such accumulation kills the animal will therefore be

$$t_{\text{death}} = \frac{C^*_{\text{poison}}}{k} \tag{4.4}$$

In this simple model, the rate of poison accumulation sets the lifespan of the animal. Any temperature dependence of the reaction 4.2 will therefore determine temperature lifespan of the animal.

We can use this model to infer the rate from lifespan–

$$k = \frac{C^*_{\text{poison}}}{t_{\text{death}}} \tag{4.5}$$

Now, C^*_{poison} is unknown and probably unmeasurable in the near future. However, we do have the ability to measure t_{death} quite well. Lifespan varies between individuals in a population, but one can summarize distributions in many ways, for example by its average $E[t]$, where $E[t] = \int_0^\infty S(t)dt$ and $S(t)$ is the survival function. Another estimate for t_{death} for populations exhibiting Weibullian mortality kinetics might be the inverse of Weibull β parameter, $\frac{1}{\beta}$. After all, $\frac{1}{beta}$ clearly represents a rate parameter for a population's logarithmic increase in hazard. In most cases this will produce results very close to mean lifespan, as the mean of a Weibull distribution is

$E[t] = \beta\Gamma(1 + 1/\alpha)$ where the complete Gamma function $\Gamma(a)$ is defined as $\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t)dt$. The function $\Gamma(1 + 1/\alpha)$ only varies between approximately 0.8 and 1 for small α and then asymptotes to 1. Thus the mean lifespan is well approximated by the Weibull β itself, as appears to be the case in our data (tables 3.2 and 4.2).

Using this definition for the time of death, we can state that

$$k = \frac{C^*_{\text{poison}}}{\beta} \tag{4.6}$$

and introducing this into the Arrhaneus equation, we obtain

$$\log \beta = \frac{E_a}{R}\frac{1}{T} + log(c) \tag{4.7}$$

Notice how the unknown parameter C^*_{poison} merges with the Arrhaneius constant A. By estimating the Weibull β at many temperatures, we can estimate E_a without knowing either the Arrhaneius A or C^*_{poison}.

We do exactly this, generating estimates for E_a, the energy of our hypothetical lifespan-limiting process, by running a linear regression to evaluate the relationship between the transformed variables $\log \beta$ and $1/T$. The results of a regression on the data collected in each temperature regime are presented in Fig. 4.7. Estimates for E_a were obtained for each temperature regime and are presented in table 4.3.

Table 4.3: Activation energies derived from fits shown in Fig. 4.7, with standard errors derived from the linear regression of transformed variables. In each case, the reaction rate was estimated as the inverse of the Weibull parameter β estimated from the first 50% of deaths. Estimates based on mean and median lifespan, as well as the Weibull β parameter including a frailty term fit on 100% of the data, produced results differing by less the standard error shown.

Temperature Range (°C)	E_a **(kJ/mol)**	E_a**(eV)**
24.8-29.1	141 ± 6	1.47 ± 0.07
30.-31.6	613 ± 43	6.36 ± 0.45
31.7-34.3	232 ± 10	2.4 ± 0.1
34.9-36.7	326 ± 19	3.37 ± 0.19

The activation energies obtained appear biologically plausible, falling in a range between 1.4 and

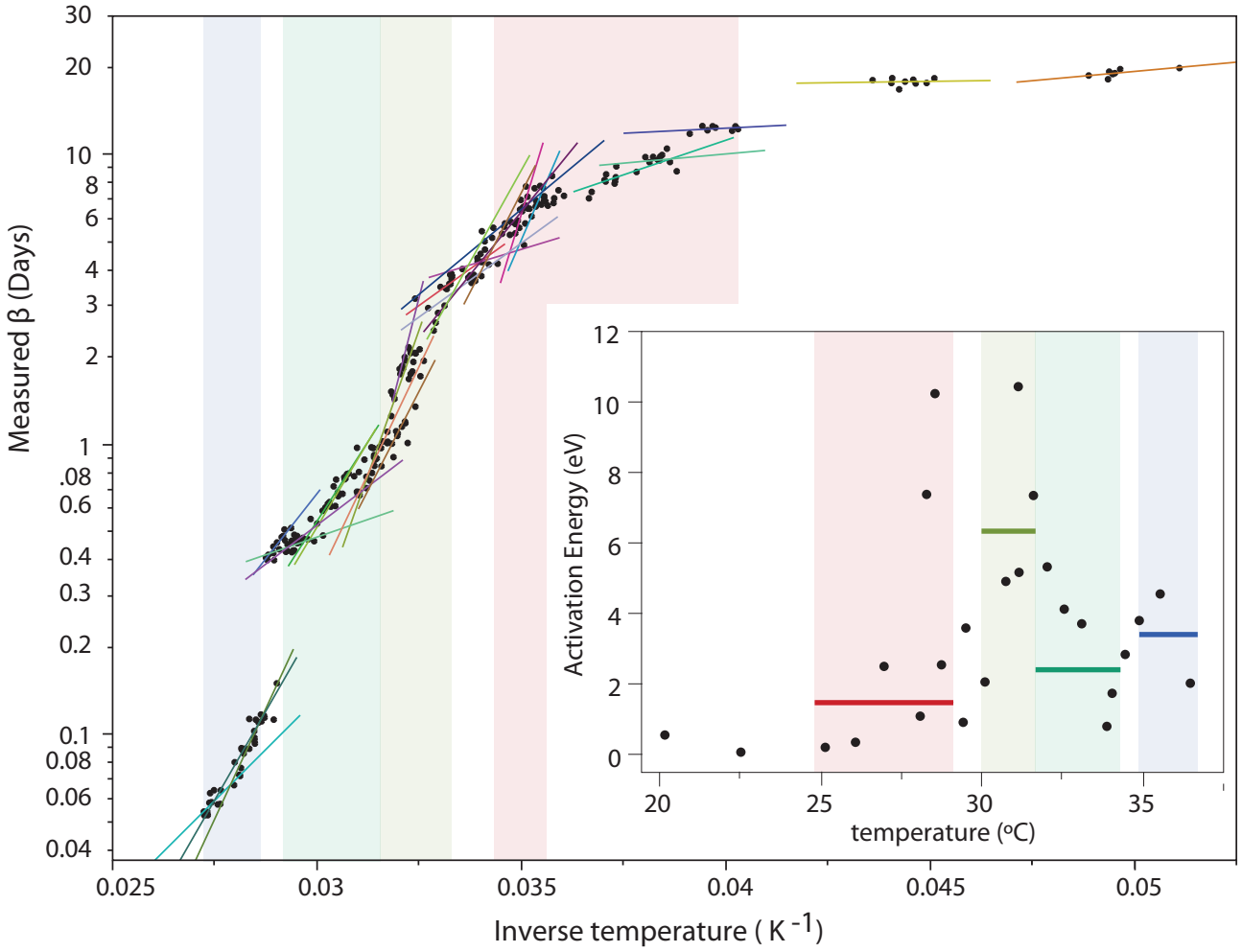

Figure 4.8: Evaluating the quality of activation energy estimates by comparing single experiment fits to trans-replicate fits Temperatures vary between scanners housed within a single incubator. This variation can be used to make rough fits of the Arrhenius equation within a single experiment *(colored lines in large figure)*. The E_as derived from each single experiment (inset, black points) are compared with the E_as obtained by pooling data from multiple experiments *(inset, colored bars)*. In both the large figure and inset, colored regions correspond to the temperature regimes analyzed in Fig. 4.7 and Table 4.3. Experiments were performed by ZN, ILM, WA, CW, and NES

7.5 electron volts (table 4.3). This is somewhat above the activation energies of single metabolic reactions, for example the Arrhenius-derived activation energy of pyruvate kinase has been estimated as 53.9 kJ/mol (0.56 eV) [Mustafa and Hochachka, 1971], and the Arrhenius-derived activation energy of amino acid incorporation by a free ribosome has been estimated as 106 kJ/mol (1.1 eV) [Towers et al., 1972].

4.6 Discussion

This chapter presented a a series of experiments where wild-type *C. elegans* were housed at 20° from hatching until their second day of adulthood at which point they were were shifted to a variety of temperatures ranging from 20 °C to 36.5°C. As expected, higher temperatures drastically reduced remaining lifespan. Surprisingly, statistical form of the distribution of remaining lifespan appeared to remain constant regardless of the temperature tested (Fig. 4.2), demonstrating a decelerating Weibullian hazard rate at timescales ranging from 14 hours to two weeks (Fig. 4.1). Deviations from a perfect scaling can be seen in the data set–for example several comparisons in the Q-R deviate from perfect horizontality 4.2. However these fall within variation observed between experimental replicates performed at the same temperature (Fig. 3.2) and therefore can reasonably be explained by limitations of our experimental method. Thus interpreted, the scaling data suggests not only that the functional form *elegans* lifespan remaining invariant to temperature, but that temperature acts primarily to set the time scale of death (the coefficient of variation (Weibull α) remaining relatively constant, 7.7 ± 0.6 at 33°C (Table 4.2) compared to 8.6 ± 0.4 °C ($\pm$ 95% confidence intervals from MLE fit)). It would therefore appear that temperature seems to act to increase the rate of aging, while having no effect on the variation in lifespan between individuals.

Several different hypothesis can be formulated to explain this phenomenon. The lifespan of each individuals might be fully specified by the second day of adulthood, which in this case would mean that the shape of the hazard curve was in our experiments a fixed initial condition. The coefficient of variation would be invariant to temperature because it was determined before any temperature change occurred. Another hypothesis would be that similar stochastic processes play out at low and high temperatures. In this case, the relative lifespan of individuals would not be determined

before the temperature shift, but would instead represent the outcome of some ongoing dynamic, stochastic process. The temperature-independence of the coefficient of variation of lifespan would therefore necessarily result from the temperature-independence of that stochastic process. This would seem most possible if variation in lifespan lifespan was not determined by a specific molecular mechanism, but was instead the result of some fundamental physical process. One candidate for this would be an extreme-value processes as discussed in 3.5. Whereas lifespan variation emerging from the biophysical behavior of specific molecules might be expected to demonstrate temperature-dependant behavior, extreme value phenomena would produce the same distribution as long as the system's architecture did not change in some fundamental way. Weibull distributions would be observed across a wide range of underlying failure distributions (3.5), even if properties of those distributions were temperature-dependent. The main constraint for EV phenomena to work in this way is that the components vital to life would need to fail independently according to the same distribution of failure times at all temperatures.

The idea that variation in wild-type *elegans* lifespan might be somehow abstracted from specific molecular mechanisms is consistent with the Arrhenius relationship observed in Fig 4.7. The existence distinct scaling regimes suggests that distinct physiological transitions occur at 30°C and 35°C, representing a change in the biochemistry causing death at low and high temperatures. Discontinuities in the scaling relationship suggest that different mechanisms determine lifespan within different temperature ranges. The shape of hazard curve appears invariant across temperature ranges between which lifespan-determining mechanisms appear to differ, suggesting that the shape of the hazard curve might be independent of such mechanisms. This does not provide formal proof of an extreme-value model underlying variation in lifespan. However, the existence a "failure architecture" in *C. elegans* would 1) explain the temperature-independence of hazard function shape and 2) explain the observed functional form of the wild type hazard curve at all temperatures.

4.6.1 Testable hypotheses

Several follow-up experiments are suggested by this data. First, it may be possible to identify the specific molecular mechanisms underlying the segmental temperature scaling of lifespan in each

temperature regime. One could imagine designing screens that would identify genes based on their effect on the Arrhenius-derived activation energies we have measured. Though many genes act to specify lifespan at any given temperature, perhaps only a subset will prove directly involved mechanistically, in a way that determines some rate of living (4.5) contributing to the activation energies we have measured. Characterization of Arrhenius phenotypes necessarily require perturbations be tested at multiple temperatures, but the population size at any given temperature need not be very large. It therefore appears experimentally possible to distinguish genes that that ”trivially” determine lifespan from those that ”fundamentally” determine lifespan, much in the same way that the *daf-16* dependence of lifespan perturbations was used to distinguish ”sick” from ”fast aging” animals in [Samuelson et al., 2007]. Short of a full unbiased screen, several candidate genes might be chosen, including for example the components of basic cell-biologic processes including mitochondrial components, metabolic enzymes, and the ribosome.

Another experimental avenue worth pursuing involves experiments seeking to differentiate between the static and dynamic models of lifespan determination. Would temperature shifts at a different point in adulthood produce results different than those observed? The experiments presented here involved the transfer of animals on their second day of adulthood. This was chosen primarily based on experimental convenience. Animals are far enough from development that their physiology has stabilized somewhat, and so the exact timing of the temperature shift within a single day of lab work might not be expected to effect experimental results. For two days, hundreds of animals can be stored on a single plate without running out of food. Beyond that point, animals must be shifted to fresh plates, significantly increasing the handling involved. Importantly, two-day old adults remain healthy enough that liquid handling might not be expected to damage them. Older animals might be significantly more fragile and require laborious, tender care. If these concerns could be addressed, the age-dependence of the temperature-dependence of lifespan might be characterized. It would be very interesting to determine, for example, whether time spent at low temperature is equivalent to that at high temperature, which might be done by quantitatively assessing the Weibull alpha experienced at high temperature by animals aged at low temperature for different amounts of time.

One final future direction concerns the Arrhenius model presented. To generate an activation energy E_a for each temperature regime, we summarized each survival curve at each temperature using quantities including the mean lifespan, the median lifespan, or the Weibull β parameter. An interesting conceptual conundrum arises here. The Weibull β is a scale term, in the sense that it describes the time scale of the hazard function. However, the conceptual basis for our Weibull fits lies with extreme value theory, where β is an emergent property determined by quantities including not only the time-scale of the underlying failure distribution for components (Fig. 3.4), but also the number of such components comprising each system (Fig. 3.5). Does the observed temperature dependence of β arise from changes to the time-scale of the underlying the failure distribution or does it act more pervasively to alter the number of components? Or, perhaps the better path for reconciling the observed variation of lifespan with Arrhenius measurements would be to take the model literally and to assume that lifespan variation corresponds directly to a distribution of activation energies E_a among individuals. Perhaps there is some biophysical expectation for distribution activation energies, derived from first principles based on the behavior of populations of independently operating biochemical systems?

Chapter 5

Reflections on the measurement of lifespan, and a few implications of the extreme-value model

Each chapter ends with its own discussion (2.3,3.7,4.6), so this final summary is relatively brief. Lifespan assays are repetitive and monotonous, with considerable effort expended by large numbers of *elegans* researchers to perform a relatively mindless task. Such are rigors of manual lifespan assays that they set up incentives to keep population sizes small and to limit the number of replicates performed. Statements on the edge of statistical significance might be justifiable within the context of individual experiments, but in aggregate contribute to a lack of reproducibility and create barriers to integrative analysis of the literature. We have developed a technique that allows researchers to perform automated assays, freeing up their time for more creative endeavors, while simultaneously allowing the observation of larger populations at a higher temporal resolution. A huge effort (2+ years) was spent ensuring that automation did not distance experimenters unnecessarily from the raw data. A major part of this involved the development of a software toolset allowing individual nematodes to be visualized in context of their late life movements. The accessibility of these visualizations introduces degree of retrospective verifiability into *elegans* lifespan, allowing assumptions on worm behavior to be evaluated in context of their quantitative effect on survival statistics. New phenomenon, for example alterations in worm behavior, gross anatomical changes in worm morphology, and (more typically) novel microbial contaminants in experiments, can be identified visually,thereby

preventing such phenomena from translating into spurious results. The author realized early on that an automated lifespan apparatus, if not sufficiently robust and self-consciously operated, could act in effect as a ”random result generator”. The high-throughput nature of such an assay would allow large numbers of assays to be performed until subsets were identified that, for entirely spurious reasons, supported a desired conclusion. The automated nature of such an assay would make it harder to perform careful experiments, and perhaps even provide plausible deniability for willful inattention to detail. This would, in the short term, generate lots of confusion and, in the long term, provide fodder for a well-deserved backlash against even careful applications of the automated method. The technologies presented here represent the author's best attempts to incorporate careful validation as a necessary step in the daily routine of automated survival assays. In this way, the technology will be more likely to improve rather than degrade the average quality of lifespan data published.

The Lifespan Machine accelerates screening efforts, allowing a wider range of biological mechanisms to be rapidly quantified, including mutants, RNAi constructs, and compounds. These types of experiments are demonstrated in chapter 2, but not followed up on in chapters 3 and 4. The reason for this is two-fold. The first is technical. The interpretation of lifespan data requires a great deal of statistical analysis, most commonly in the application of test statistics used to evaluate equality of two lifespan distributions. The basic assumptions of many statistical approaches used to evaluate the significance of lifespan perturbations remain relatively untested for *C. elegans* lifespan data. Do mutations act to alter a population's hazard by a time-invariant proportional constant? If this was not the case, then a wide range of test statistics will provide, at best, only approximate estimates of statistical significance. When considering dramatic lifespan perturbations such as the 100% extension of *daf-2(e1370)*, these nuances are relatively inconsequential. However, for smaller perturbations, deviations from model assumptions entirely determine the behavior of test statistics. Proper extraction of biologically-relevant conclusions from high-resolution *elegans* survival curves requires an improved understanding of the underlying statistical properties of emperical distributions of *elegans* death times. Clarification of the functional form of wildtype *elegans* serves as an important basis for the evaluation of various statistical tests used to evaluate perturbations to lifespan. The close fit of a

Weibull model incorporating a frailty term to the deaths of wild type *elegans* gives a strong prediction for the strengths and weaknesses of various tests, at least when considering perturbations that act as temperature does to affect primarily the Weibull β.

The distributional form of *elegans* death times is also itself intrinsically interesting. The shape of the distribution of death times represents the statistical signature of the underlying molecular processes determining lifespan. Initial characterization of high resolution lifespan data produced a very fortuitous result–the shape of the *C. elegans* hazard function appears surprisingly simple, even when examined in (according to current standards) very large populations. Deaths are well described by a simple Weibull model with a frailty term added to account for heterogeneity within the population. Future work may reveal this apparent simplicity to be an ironic reflection of our current inability to collect experimental data of sufficient quality. Or, the apparent simplicity might result from our failure to apply sufficiently powerful analytic to identify the complexity existing in our data. However, where simple distributional forms arise from complex phenomena, for example as is the case in the black-body spectrum of the cosmic microwave background, one begins to suspect some basic principle may be acting to unify disparate phenomenon. Much effort has been put into inferring such principles from mortality data, starting with Gompertz in the nineteenth century ([Gompertz, 1825]) and continuing throughout the 20th ([Strehler and Mildvan, 1960, Steinsaltz, 2005, Wu et al., 2006, Gavrilov and Gavrilova, 2001]). Much of this conceptual work, however, focuses on the Gompertzian form of lifespan which appears to be a good model for death in human populations. Recent work in model organisms subsequently imported the Gompertzian model as framework interpreting experimental investigations. We find that the Weibull distribution is as good if not better at describing the mortality of *elegans*, necissitating the development of new models for *elegans* aging.

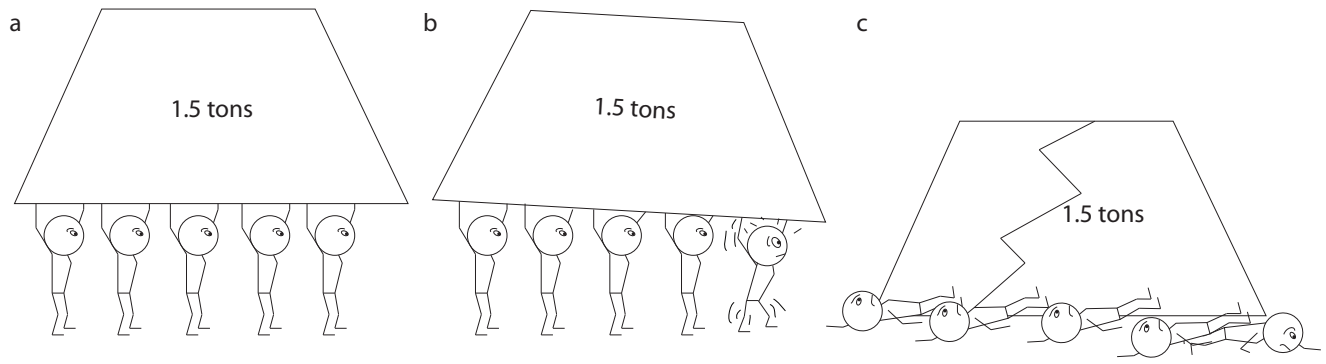

Figure 5.1: A physical model of an extreme-value phenomena. (a) Several individuals cooperate to accomplish a task none could handle alone. In fact, the task requires the combined efforts of all individuals, in such a way that **(b)**the failure of the first person **(c)** causes the failure of the group

Our data concerning the statistical form of mortality distributions at low (Fig 3.2) and high (Fig 4.1) temperatures suggest that *elegans* die according to a Weibull distribution, and we interpret the temperature scaling data (Fig 4.2) collected to suggest that this distributional form may arise independantly of any specific molecular mechanism. We therefore propose a model whereby the distribution of death times is rendered independent from the distribution of failure times of biological components by virtue of some extreme value phenomenon. Clearly, the ongoing survival of an animal depends on the continued interaction of various biological components. We present no direct evidence as to the identity of such components–and each ”component” might itself be divisible. A component could be a cell, a class of molecules, a physiological property (the pH or redox state of cells) or a tissue. Instead, we present model that describes such components in relation to their interaction's specification failure time of an individual. Succinctly, our data is consistent with *elegans* being composed of a set of vital components, the failure of any single component killing the animal (Fig. 5.1). Such components would need to fail independently according to the same distribution of failure times. The effect of this type of mutually truncating failure process is to mask the distribution of failure times of any individual component. The distribution of the first failure of components will converge to a Weibull distribution for a wide range of underlying failure distributions 3.5. This appears to provide a simple explanation for the apparent independence of survival curve shape and molecular mechanism.

This type of failure architecture might explain why the molecular determinants of lifespan lifespan remain elusive. The failure of any single vital component might suffice to kill the entire animal, subsequently breaking all components (Fig. 5.2). Even if the cause of death was definitively identified in a single animal, subsequent replicates would show the fist failure happening in different components, and at different times. Furthermore, the first failure might as a consequence lead to the failure of other components, creating the situation where all components appeared to degrade in tandem.5.2. The death of each animal would have a single cause but the causative mechanism will differ between individuals, occur at different times in different individuals, and be masked by the consequential failure of other vital mechanisms.

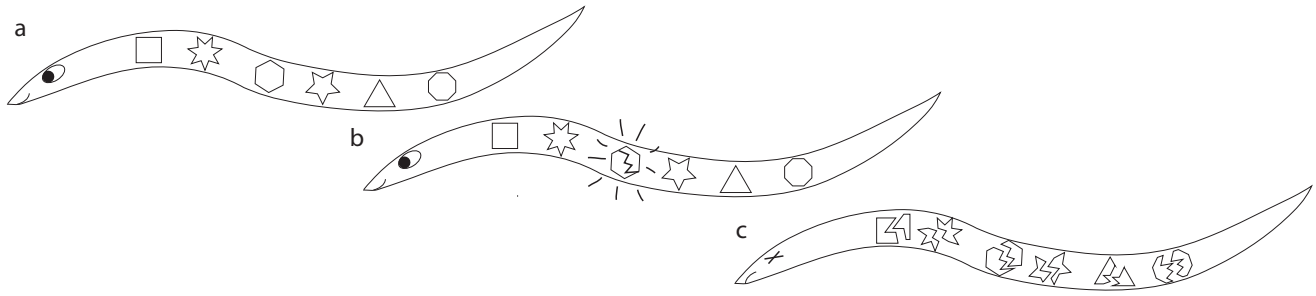

Figure 5.2: The lifespan of an organism, truncated at the time of the first failure among its vital components (a) An animal begins with all its components intact. **(b)** Eventually, a component fails, **(c)** killing the animal and breaking all the other components.

The biggest constraint on our model of the failure architecture of *C. elegans* is that Weibull distributions arise in theory only from systems comprised of components with identical failure distributions. Much more work needs to be done to identify the quantitative details of such a constraint–one imagine similar but not identical failure distributions might combine to produce distributions that approximate a Weibull to the same extent as our experimental data. One can speculate that evolutionary processes might produce systems organized in this way. Evolutionary selection acts against components that decrease the fitness of the individual. Clearly, the spontaneous failure of a vital component (from wear and tear, by misadventure, etc) would, early in life, kill the animal and lower its fitness. One solution would be for organisms to develop mechanisms to ensure robustness, in effect spreading important functions across a sufficient

number of parts that a failure in any single part would have no dramatic effect. However, such a design necessarily must be limited by the reality of metazoan tissues. A set of cells might be designated to all perform the same task, allowing any single cell to die without consequence. However, metazoans accomplish an array of important tasks through specialization of cell types. The talents of different cell types may overlap, there are limits to how much one cell type can compensate for another. This would imply that completely distributed robustness may be impossible to realize in practice–a body wall muscle cell will likely be unable compensate for the failure of an excretory cell. These barriers to interoperability of cell-type would then cause related cells to act in an effect as single component, whose failure kills the organism. If all parts of an animal could compensate for all other parts, then such robustness might fundamentally change the failure dynamics of the system. However, as long as at some level of biological organization, components are not interchangeable, Extreme value phenomena will arise. In such a system selective pressures will then be likely to to balance the robustness of individual components in such a way that, though they necessarily failed independently, they did so with roughly the same statistical dynamics. Evolution might create animals without ”a weakest link”.

One might further speculate that such a coordination of failure distributions (robustness) of individual tissues might be sufficiently consequential at short time-scales that they were placed under neuro-endocrine control. The physiological state of cells and tissues might be monitored and their failure anticipated, so that weaknesses could be made more robust within the lifespan of a single animal. For example, if an individual realized that its gut lining was about to fail in such a way that would truncate the individual's lifespan, it might want to quickly increase the robustness of its gut, perhaps by expressing of chaperones or altering metabolic processes. This type of ”just-in-time” tissue maintenance would continue independently in each tissue until it failed in one, killing the individual. Furthermore, if animals encountered an environment where it was advantageous to alter their average death time–for example to outlive a period of famine or stress–such independent maintenance of each tissue type would benefit from whole-animal coordination, where a vast array of underlying mechanisms altered in concert. This would suggest that evolutionary constraints on the failure architecture of *elegans* might effect the organization of lifespan and stress resistance regulatory mechanisms, and might ultimately be used explain

some part of the fascinating complexity observed in the tissue-specific action of insulin/IGF signalling pathway components ([Arantes-Oliveira et al., 2002, Apfeld and Kenyon, 1998].

Bibliography

[Aalen et al., 2008] Aalen, O., Borgan, O., and Gjessing, H. (2008). *Survival and Event History Analysis : A Process Point of View*. New York, NY : Springer New York.

[Ailion et al., 1999] Ailion, M., Inoue, T., Weaver, C. I., Holdcraft, R. W., and Thomas, J. H. (1999). Neurosecretory control of aging in Caenorhabditis elegans. *Proc Natl Acad Sci U S A 96*, 15:7394–7397.

[Amrit and May, 2010] Amrit, F.R.and Boehnisch, C. and May, R. (2010). Phenotypic covariance of longevity, immunity and stress resistance in the caenorhabditis nematodes. *PLoS One*, 5:e9978.

[Apfeld and Kenyon, 1998] Apfeld, J. and Kenyon, C. (1998). Cell nonautonomy of c. elegans daf-2 function in the regulation of diapause and life span. *Cell*, 95(2).

[Apfeld et al., 2004] Apfeld, J., O'Connor, G., McDonagh, T., DiStefano, P. S., and R, C. (2004). The AMP-activated protein kinase AAK-2 links energy levels and insulin-like signals to lifespan in C. elegans. *Genes Dev*, 18(24):3004–3009.

[Arantes-Oliveira et al., 2002] Arantes-Oliveira, N., Apfeld, J., Dillin, A., and Kenyon, C. (2002). Regulation of life-span by germ-line stem cells in caenorhabditis elegans. *Science*, 295:502–5.

[Atkins, 2001] Atkins, P. (2001). *Atkins' Physical Chemistry*. Oxford Higher Education.

[Austad et al., 2007] Austad, S., Kirkwood, T., and Guarente, L. P. (2007). *Molecular Biology of Aging*. Cold Spring Harbor Laboratory Press.

[Baeriswyl et al., 2009] Baeriswyl, S., Diard, M., Mosser, T., Leroy, M., Manière, X., Taddei, F., and Matic, I. (2009). Modulation of aging profiles in isogenic populations of Caenorhabditis elegans by bacteria causing different extrinsic mortality rates. *Biogerontology*, 11:53–65.

[Ben-Zvi A, 2009] Ben-Zvi A, Miller EA, M. R. (2009). Collapse of proteostasis represents an early molecular event in caenorhabditis elegans aging. *Proc Natl Acad Sci*, 106(35):14914–9.

[Boaro et al., 1998] Boaro, S., Soares, J., and K□nig, B. J. (1998). Comparative structural analysis of neuromuscular junctions in mice at different ages. *Ann Anat*, 180(2).

[Boser et al., 1992] Boser, B. E., Guyon, I., and Vapnik, V. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 144–152. ACM Press.

[Buckley and James, 1979] Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, 66(3):429–436.

[Budovskaya et al., 2008] Budovskaya, Y., Wu, K., Southworth, L., Jiang, M., Tedesco, P., Johnson, T., and Kim, S. (2008). An elt-3/elt-5/elt-6 GATA transcription circuit guides aging in c. elegans. *Cell*, 134(2).

[Burnham and Anderson, 2004] Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference: Understanding aic and bic in model selection. *Sociological Methods and Research*, 33(2):261–304.

[Carey et al., 1995] Carey, J. R., Liedo, P., and Vaupel, J. W. (1995). Mortality dynamics of density in the Mediterranean fruit fly. *Exp Gerontol*, 30:605–629.

[Chaui-Berlinck et al., 2004] Chaui-Berlinck, J., Navas, C., Monteiro, L., and Bicudo, J. (2004). Temperature effects on a whole metabolic reaction cannot be inferred from its components. *Proc Biol Sci*, 271(1546).

[Clark, 1991] Clark, W. (1991). Analysis of reliability data for mechanical systems. In *Reliability and Maintainability Symposium, 1991. Proceedings., Annual*, pages 438–441.

[Conti et al., 2006] Conti, B., Sanchez-Alavez, M., Winsky-Sommerer, R., Morale, M., Lucero, J., Brownell, S., Fabre, V., Huitron-Resendiz, S., Henriksen, S., Zorrilla, E., de Lecea, L., and Bartfai, T. (2006). Transgenic mice with a reduced core body temperature have an increased life span. *Science*, 314:825–8.

[Cox, 2008] Cox, C. (2008). The generalized F distribution: an umbrella for parametric survival analysis. *Statist Med*, 27(21):4301–4312.

[Curtsinger et al., 1992] Curtsinger, J. W., Fukui, H. H., Townsend, D. R., and Vaupel, J. W. (1992). Demography of genotypes: failure of the limited life-span paradigm in Drosophila melanogaster. *Science*, 258:461–463.

[David et al., 2010] David, D. C., Ollikainen, N., Trinidad, J. C., Cary, M. P., Burlingame, A. L., and Kenyon, C. (2010). Widespread protein aggregation as an inherent part of aging in <italic>c. elegans</italic>. *PLoS Biol*, 8(8):e1000450.

[Dillin et al., 2002] Dillin, A., Crawford, D. K., and Kenyon, C. (2002). Timing requirements for insulin/igf-1 signaling in c elegans. *Science*, 298:830–834.

[Dolle et al., 2000] Dolle, M., Snyder, W., Gossen, J., Lohman, P., and Vijg, J. (2000). Distinct spectra of somatic mutations accumulated with age in mouse heart and small intestine. *Proc Natl Acad*, 97(15):8403–8.

[Dong and Suen, 2005] Dong, J.X.and Krzyzak, A. and Suen, C. (2005). Fast SVM training algorithm with decomposition on very large data sets. *IEEE Trans Pattern Anal Mach Intell*, 27:603–618.

[Duchateau, 2008] Duchateau, L. J. P. (2008). *The Frailty Model*. New York, NY: Springer Science.

[Fisher and Tippett, 1928] Fisher, R. A. and Tippett, L. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24:180–190.

[Friedman and Johnson, 1986] Friedman, D. and Johnson, T. (1986). Further genetic analysis of a long-life mutant strain. *C. elegans Newsletter*, 9.(1):99.

[Fuster et al., 1992] Fuster, V., Badimon, L., Badimon, J. J., and Chesebro, J. H. (1992). The pathogenesis of coronary artery disease and the acute coronary syndromes. *New England Journal of Medicine*, 326(4):242–250.

[Gavrilov and Gavrilova, 2001] Gavrilov, L. A. and Gavrilova, N. S. (2001). The reliability theory of aging and longevity. *Journal of Theoretical Biology*, 213(4):527–545.

[Gems and Riddle, 2000] Gems, D. and Riddle, D. L. (2000). Genetic, behavioral and environmental determinants of male longevity in Caenorhabditis elegans. *Genetics*, 154(4):1597–1610.

[Gompertz, 1825] Gompertz, B. (1825). On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London*, 115:513–585.

[Hamilton et al., 2005] Hamilton, B., Dong, Y., Shindo, M., Liu, W., Odell, I., Ruvkun, G., and Lee., S. S. (2005). A systematic RNAi screen for longevity genes in C. elegans. *Genes Dev.*, 19(13):1544–1555.

[Hansen et al., 2005] Hansen, M., Hsu, A.-L., Dillin, A., and Kenyon, C. (2005). New Genes Tied to Endocrine, Metabolic, and Dietary Regulation of Lifespan from a Caenorhabditis elegans Genomic RNAi Screen. *PLoS Genet*, 1(1):e17.

[Haskins et al., 2008] Haskins, K.A., R., J.F., Gaddis, N., Dressman, H., and Aballay, A. (2008). Unfolded protein response genes regulated by CED-1 are required for Caenorhabditis elegans innate immunity. *Dev Cell*, 15:87–97.

[Herndon et al., 2002] Herndon, L. A., Schmeissner, P. J., Dudaronek, J. M., Brown, P. A., Listner, K. M., Sakano, Y., Paupard, M. C., Hall, D. H., and Driscoll, M. (2002). Stochastic and genetic factors influence tissue-specific decline in ageing C. elegans. *Nature*, 419(6909):808–814.

[Heron, 2012] Heron, M. (2012). Deaths: Leading Causes for 2009. *National Vital Statistics Reports*, 61(7).

[Hess et al., 1999] Hess, K. R., Serachitopol, D. M., and Brown, B. W. (1999). Hazard function estimators: a simulation study. *Statistics in Medicine*, 18(22):3075–3088.

[Hosono, 1978] Hosono, R. (1978). Sterilization and growth inhibition of Caenorhabditis elegans by 5-fluorodeoxyuridine. *Experimental Gerontology*, 13(5):369–374.

[Huang et al., 2004] Huang, C., Xiong, C., and Kornfeld, K. (2004). Measurements of age-related changes of physiological processes that predict lifespan of Caenorhabditis elegans. *Proc Natl Acad Sci U S A*, 101:8084–8089.

[Isied et al., 1984] Isied, S. S., Kuehn, C., and Worosila, G. (1984). Ruthenium-modified cytochrome c: temperature dependence of the rate of intramolecular electron transfer. *J. Am. Chem. Soc.*, 106(6):1722?1726.

[Jackson et al., 2010] Jackson, C., Sharples, L., and Thompson, S. (2010). Survival models in health eco nomic evaluations: balancing fit and parsimony to improve prediction. *International Journal of Biostatistics*, 6(1).

[Johnson et al., 2001] Johnson, T. E., Wu, D., Tedesco, P., Dames, S., and Vaupel, J. W. (2001). Age-specific demographic profiles of longevity mutants in Caenorhabditis elegans show segmental effects. *J Gerontol A (Biol Sci Med Sci)*, 56:B331–339.

[Jones, 2008] Jones, J. (2008). *GIMP User's Manual*.

[Kaplan and Meier, 1958a] Kaplan, E. and Meier, P. (1958a). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53(282):457–481.

[Kaplan and Meier, 1958b] Kaplan, E. L. and Meier, P. (1958b). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, pages 457–481.

[Kenyon, 2010] Kenyon, C. (2010). The genetics of ageing. *Nature*, 464:504–512.

[Kenyon et al., 1993] Kenyon, C., Chang, J., Gensch, E., Rudner, A., and Tabtiang, R. (1993). A C. elegans mutant that lives twice as long as wild type. *Nature*, 366:461–464.

[Kimura et al., 1997] Kimura, K. D., Tissenbaum, H. A., Liu, Y., and Ruvkun, G. (1997). daf-2, an insulin receptor-like gene that regulates longevity and diapause in Caenorhabditis elegans. *Science*, 277:942–946.

[Kiontke et al., 2011] Kiontke, K., Felix, M., Ailion, M., Rockman, M., Braendle, C., Penigault, J., and Fitch, D. (2011). A phylogeny and molecular barcodes for Caenorhabditis, with numerous new species from rotting fruits. *BMC Evol Biol*, 339(11).

[Kirkwood et al., 2005] Kirkwood, T., Feder, M., Finch, C., Franceschi, C., Globerson, A., Klingenberg, C., LaMarco, K., Omholt, S., and Westendorp, R. (2005). What accounts for the wide variation in life span of genetically identical organisms reared in a constant environment. *Mech Ageing Dev*, 126(3):439–443.

[Klass, 1977] Klass, M. (1977). Aging in the nematode Caenorhabditis elegans: major biological and environmental factors influencing life span. *Mech Ageing Dev*, 6:413–429.

[Kuhn, 1955] Kuhn, H. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 53.

[Laidler, 1984] Laidler, K. J. (1984). The development of the arrhenius equation. *Journal of Chemical Education*, 61(6).

[Larsen et al., 1995] Larsen, P. L., Albert, P. S., and Riddle, D. L. (1995). Genes that regulate both development and longevity in Caenorhabditis elegans. *Genetics*, 139:1567–1583.

[Lee, 2006] Lee, S. (2006). Whole genome rnai screens for increased longevity: important new insights but not the whole story. *Exp Gerontol.*, 41(10).

[Lee et al., 2002] Lee, S. S., Lee, R. Y., Fraser, A. G., Kamath, R. S., Ahringer, J., and Ruvkun, G. (2002). A systematic RNAi screen identifies a critical role for mitochondria in C. elegans longevity. *Nature Genetics*, 33(40 - 48).

[Lin et al., 1997] Lin, K., Dorman, J., Rodan, A., and Kenyon, C. (1997). daf-16: An HNF-3/forkhead family member that can function to double the life-span of Caenorhabditis elegans. *Science*.

[Lithgow et al., 1995] Lithgow, G., White, T., Melov, S., and Johnson, T. ((1995)). Thermotolerance and extended life-span conferred by single-gene mutations and induced by thermal stress. *Proc Natl Acad Sci U S A*, 92:7540–7544.

[Loram and Bodnar, 2012] Loram, J. and Bodnar, A. (2012). Age-related changes in gene expression in tissues of the sea urchin Strongylocentrotus purpuratus. *Mechanisms of Ageing and Development*, 133(5).

[Maduro et al., 2000] Maduro, M. F., Gordon, M., Jacobs, R., and Pilgrim, D. B. (2000). The UNC-119 family of neural proteins is functionally conserved between humans, Drosophila and C. elegans. *Journal of Neurogenetics*, 13(4):191–212.

[Maegawa et al., 2010] Maegawa, S., Hinkal, G., Kim, H., Shen, L., Zhang, L., Zhang, J., Zhang, N., Liang, S., Donehower, L., and Issa, J. (2010). Widespread and tissue specific age-related dna methylation changes in mice. *Genome Res.*, 20(3).

[Mair et al., 2003] Mair, W., Goymer, P., Pletcher, S., and Partridge, L. (2003). Demography of dietary restriction and death in Drosophila. *Science*, 301:1731–1733.

[Mathew et al., 2012] Mathew, M., Mathew, N., and Ebert, P. (2012). WormScan: A Technique for High-Throughput Phenotypic Analysis of Caenorhabditis elegans. *PLoS ONE*, 7(3):e33483.

[McColl et al., 2010] McColl, G., Rogers, A., Alavez, S., Hubbard, A., Melov, S., Link, C., Bush, A., Kapahi, P., and Lithgow, G. (2010). Insulin-like signaling determines survival during stress via posttranscriptional mechanisms in C. elegans. *Cell Metab.*, 12(3):260–72.

[McGee et al., 2011] McGee, M. D., Weber, D., Day, N., Vitelli, C., Crippen, D., Herndon, L. A., Hall, D. H., and Melov, S. (2011). Loss of intestinal nuclei and intestinal integrity in aging C. elegans. . *Aging Cell*, 10(4):699–710.

[Miller and Austad, 1999] Miller, R. and Austad, S. (1999). Large animals in the fast lane. *Science*, 285(5425):199.

[Miquel et al., 1976] Miquel, J., Lundgren, P. R., Bensch, K. G., and Atlan, H. (1976). Effects of temperature on the life span, vitality, and fine structure of drosophila melanogaster. *Mechanisms of Ageing and Development*, 5:347–370.

[Morley and Morimoto, 2004] Morley, J. F. and Morimoto, R. I. (2004). Regulation of longevity in Caenorhabditis elegans by heat shock factor and molecular chaperones. *Mol Biol Cell*, 15(2):657–664.

[Mueller and Wang, 1994] Mueller, H. and Wang, J. (1994). Hazard rates estimation under random censoring with varying kernels and bandwidths. *Biometrics*, 50:61–76.

[Murphy, 2005] Murphy, C. T. (2005). A Review of Genes that Act Downstream of the DAF-16 FOXO transcription factor to influence the lifespan of C. elegans. *Research and Perspectives in Longevity, Longevity and Frailty,Springer*.

[Mustafa and Hochachka, 1971] Mustafa, T. and Hochachka, P. (1971). Catalytic and regulatory properties of pyruvate kinases in tissues of a marine bivalve. *J Biol Chem*, 246(10).

[Oh et al., 2005] Oh, S., Mukhopadhyay, A., Dixit, B., Raha, T., Green, M., and Tissenbaum, H. (2005). Identification of direct daf-16 targets controlling longevity, metabolism and diapause by chromatin immunoprecipitation. *Nature Genetics*, 38(2):251–7.

[Pearl, 1928] Pearl, R. (1928). *The Rate of Living*. Ulan Press (2012 reprint).

[Pincus et al., 2011] Pincus, Z., Smith-Vikos, T., and FJ, S. (2011). MicroRNA predictors of longevity in Caenorhabditis elegans. *PLoS Genet*, page e1002306.

[R Development Core Team, 2009] R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[Raamsdonk and Hekimi, 2012] Raamsdonk, J. M. V. and Hekimi, S. (2012). Superoxide dismutase is dispensable for normal animal lifespan. *PNAS*, 109(15):5785?5790.

[Rea et al., 2005] Rea, S. L., Wu, D., Cypser, J. R., Vaupel, J. W., and Johnson, T. E. (2005). A stress-sensitive reporter predicts longevity in isogenic populations of caenorhabditis elegans. *Nat Genet 37*, 47:894–898.

[Reinsch, 1967] Reinsch, C. H. (1967). Smoothing by Spline Functions. *Numerische Mathematik*, 10(3):177.

[Rinne, 2008] Rinne, H. (2008). *The Weibull Distribution: A Handbook*. Chapman and Hall/CRC.

[Rose et al., 2008] Rose, M., Burke, M., Shahrestani, P., and Mueller, L. (2008). Evolution of ageing since darwin. *J. Genet*, 87(4).

[Ruggiero and Ferrucci, 2006] Ruggiero, C. and Ferrucci, L. (2006). The Endeavor of High Maintenance Homeostasis: Resting Metabolic Rate and the Legacy of Longevity. *J Gerontol A Biol Sci Med Sci.*, 61(5).

[Sall et al., 2012] Sall, J., Lehman, A., and Stephens, M. (2012). *JMP Start Statistics: a Guide to Statistics and Data Analysis Using Jmp, Fifth Edition*. SAS Institute.

[Samuelson et al., 2007] Samuelson, A., Carr, C., and Ruvkun, G. (2007). Gene activities that mediate increased life span of c. elegans insulin- like signaling mutants. *Genes Dev*, 21(22):2976–94.

[Sekiya et al., 2009] Sekiya, M., Nakamoto, R. K., Al-Shawi, M. K., Nakanishi-Matsui, M., , and Futai, M. (2009). Temperature dependence of single molecule rotation of the escherichia coli atp synthase f1 sector reveals the importance of ?-? subunit interactions in the catalytic dwell. *J Biol Chem.*, 284(33).

[Sethna, 2006] Sethna, J. P. (2006). *Statistical Mechanics: Entropy, Order Parameters and Complexity*. Oxford University Press, USA.

[Singer and Willett, 2003] Singer, J. and Willett, J. (2003). *Applied Longitudinal Data Analysis*. Oxford University Press.

[Somero and Hochachka, 1968] Somero, G. and Hochachka, P. (1968). The effect of temperature on catalytic and regulatory functions of pyruvate kinases of the rainbow trout and the antarctic fish trematomus bernacchii. *Biochem J.*, 110(3):395–400.

[SP and G., 2007] SP, C. and G., R. (2007). Lifespan regulation by evolutionarily conserved genes essential for viability. *PLoS Genet*, 3(4):e56.

[Stata Corporation, 2007] Stata Corporation (2007). *Stata Base Reference Manual : Release 10*. College Station, TX : Stata Press.

[Steinsaltz, 2005] Steinsaltz, D. (2005). Re-evaluating a test of the heterogeneity explanation for mortality plateaus. *Exp Gerontol.*, 40(1-2):101–13.

[Steinsaltz and Evans, 2004] Steinsaltz, D. and Evans, S. N. (2004). Markov mortality models: implications of quasistationarity and varying initial distributions. *Theoretical Population Biology*, 65(4):319–337.

[Steinsaltz and Evans, 2007] Steinsaltz, D. and Evans, S. N. (2007). Quasistationary distributions for one-dimensional diffusions with killing. *Transactions of the American Mathematical Society*, 359(3):1285–1324.

[Stiernagle, 2006] Stiernagle, T. (2006). *Maintenance of C elegans*. WormBook.

[Strehler and Mildvan, 1960] Strehler, B. and Mildvan, A. (1960). General Theory of Mortality and Aging. *Science*, 132(3418):14–21.

[Suda et al., 2012] Suda, H., Sato, K., and Yanase, S. (2012). Timing mechanism and effective activation energy concerned with aging and lifespan in the long-lived and thermosensory mutants of caenorhabditis elegans. *Mechanisms of Ageing and Development*, 133:600–610.

[Sutphin and Kaeberlein, 2009] Sutphin, G. and Kaeberlein, M. (2009). Measuring Caenorhabditis elegans life span on solid media. *J Vis Exp*.

[Taguchi et al., 2007] Taguchi, A., Wartschow, L. M., and White, M. F. (2007). Brain irs2 signaling coordinates life span and nutrient homeostasis. *Science*, 317.

[Tatar1 et al., 2001] Tatar1, M., Kopelman, A., Epstein, D., Tu, M.-P., Yin, C.-M., and Garofalo, R. S. (2001). A mutant drosophila insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science*, 292(5514):107–110.

[Towers et al., 1972] Towers, N., Raison, J., Kellerman, G., and Linnane, A. (1972). Effects of temperature-induced phase changes in membranes on protein synthesis by bound ribosomes. *Biochim Biophys Acta.*, 287(2).

[Tullet et al., 2008] Tullet, J. M., Hertweck, M., An, J., Baker, J., Hwang, J., Liu, S., Oliveira, R., Baumeister, R., and Blackwell, T. (2008). Direct inhibition of the longevity-promoting factor SKN-1 by insulin-like signaling in c elegans. *Cell*, 132:1025–1038.

[Turnbull, 1976] Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B*, 38(3):290–295.

[Vanfleteren et al., 1998] Vanfleteren, J., De Vreese, A., and Braeckman, B. (1998). Two-parameter logistic and Weibull equations provide better fits to survival data from isogenic populations of Caenorhabditis elegans in axenic culture than does the Gompertz model. *J Gerontol A Biol Sci Med Sci*, 53:B393–403; discussion B404–398.

[Vaupel et al., 1998] Vaupel, J. W., Carey, J. R., Christensen, K., Johnson, T. E., Yashin, A. I., Holm, N. V., Iachine, I. A., Kannisto, V., Khazaeli, A. A., Liedo, P., Longo, V. D., Zeng, Y., Manton, K. G., and Curtsinger, J. W. (1998). Biodemographic trajectories of longevity. *Science*, 280:855–860.

[Walker et al., 2005] Walker, G., Houthoofd, K., Vanfleteren, J., and Gems, D. (2005). Dietary restriction in c elegans: from rate-of-living effects to nutrient sensing pathways. *Mechanisms of Ageing and Development*, 126(9):929–37.

[Wang et al., 1998] Wang, J.-L., Muller, H.-G., and Capra, W. (1998). Analysis of oldest-old mortailty: Lifetables revisited. *The Annals of Statistics*, 26(1):126–163.

[Ward et al., 2008] Ward, A., Liu, J., Feng, Z., and Xu, X. Z. S. (2008). Light-sensitive neurons and channels mediate phototaxis in C. elegans. *Nature Neuroscience*, 11(8):916–922.

[Weitz and Fraser, 2001a] Weitz, J. and Fraser, H. (2001a). Explaining mortality rate plateaus. *Proc Natl Acad Sci U S A*, 98:15383–15386.

[Weitz and Fraser, 2001b] Weitz, J. and Fraser, H. (2001b). Explaining mortality rate plateaus. *Proc Natl Acad Sci USA*, 98(26):15383–15386.

[Wilk and Gnanadesikan, 1968] Wilk, M. B. and Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*, 55(1):pp. 1–17.

[Wilkinson et al., 2012] Wilkinson, D., Taylor, R., and A., D. (2012). Analysis of aging in Caenorhabditis elegans. *Methods Cell Biol*, 107:353–381.

[Wu et al., 2009] Wu, D., Rea, S., Cypser, J., and Johnson, T. (2009). Mortality shifts in Caenorhabditis elegans: remembrance of conditions past. *Aging Cell*, 8:666–675.

[Wu et al., 2006] Wu, D., Rea, S. L., Yashin, A. I., and Johnson, T. E. (2006). Visualizing hidden heterogeneity in isogenic populations of c. elegans. *Experimental Gerontology*, 41:261–270.

[Wuhr et al., 2011] Wuhr, M., Obholzer, N., Megason, S., Detrich, H. r., and Mitchison, T. (2011). Live imaging of the cytoskeleton in early cleavage-stage zebrafish embryos. *Methods Cell Biol*, 101:1–18.

[Zahn et al., 2007] Zahn, J., Poosala, S., Owen, A., Ingram, D., Lustig, A., Carter, A., Weeraratna, A., Taub, D., Gorospe, M., Mazan-Mamczarz, K., Lakatta, E., Boheler, K., Xu, X., Mattson, M., Falco, G., Ko, M., Schlessinger, D., Firman, J., Kummerfeld, S., Wood, W. r., Zonderman, A., Kim, S., and Becker, K. (2007). Agemap: a gene expression database for aging in mice. *PLoS Genet*, 3(11).

[Zhang et al., 2009] Zhang, Y., Ikeno, Y., Qi, W., Chaudhuri, A., Li, Y., Bokov, A., Thorpe, S., Baynes, J., Epstein, C., Richardson, A., and Van Remmen, H. (2009). Mice deficient in both mn superoxide dismutase and glutathione peroxidase-1 have increased oxidative damage and a greater incidence of pathology but no reduction in longevity. *J Gerontol A Biol Sci Med Sci.*, 64(12):1212–20.

Supplementary Notes and Figures

The text and figures for this supplement are taken from the supplementary materials of the manuscript entitled ”The *C. elegans* lifespan machine”, presented in chapter 2. A version of the manuscript lacking Fig. 2.5 is currently in press at the journal Nature Methods. The authors of this manuscript are Nicholas Stroustrup, Bryne E. Ulmschneider, Zachary M. Nash, Isaac F. López Moyado, Javier Apfeld, and Walter Fontana. The author contributions are as follows—N.S. designed and implemented hardware and software. N.S. and B.U. constructed and calibrated equipment. N.S. and J.A. conceived and designed experiments. N.S., B.U., J.A., Z.N. and I.F.L.M. performed experiments. N.S. designed analytic tools. N.S., J.A. and W.F. provided guidance, analyzed data, interpreted results, and wrote the manuscript. J.A and W.F. are co-last authors.

Supplementary Note S1 — Lifespan Machine hardware

S1.1 Image acquisition

Our reference installation consists of 50 scanners allocated to 5 incubators. A single scanner is sufficient to obtain controlled survival curves of high accuracy, while a fleet of scanners enables high throughput. In Note S15, we provide a description of use cases for varying numbers of scanners. Instructions for the modification of a commercially available scanner model (Epson v700) are provided as a separate supplementary document. At www.lifespanmachine.org, we have set up pointers to a public repository for the software components of the Lifespan Machine to facilitate user-driven development and improvement.

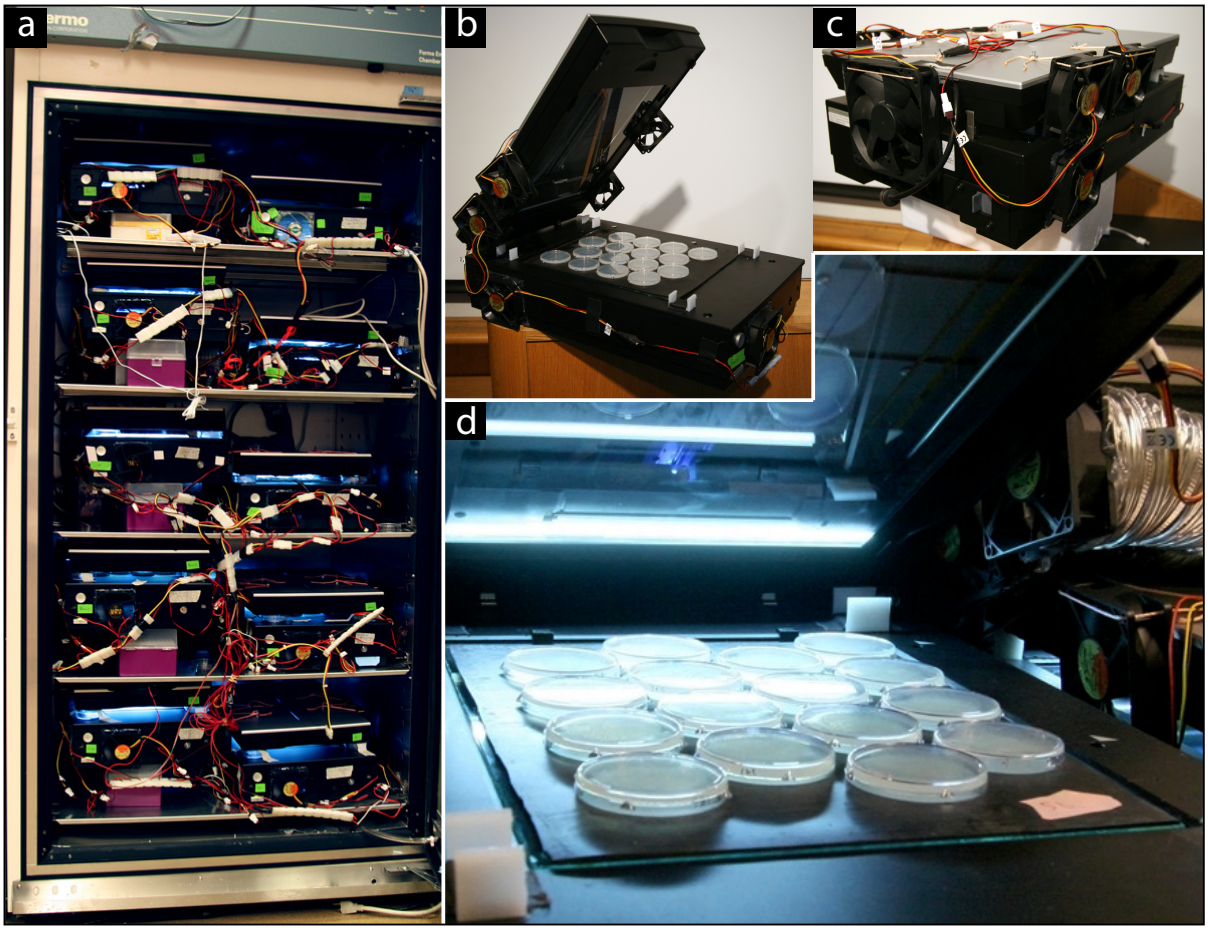

Figure S1-1: Lifespan Machine hardware. **(a)** A Thermo Forma incubator provides temperature control for ten flatbed scanners. **(b)** A scanner unit with open lid and plates loaded, showing fans mounted on the lid to create a flow of air between lid and scanner surface when the lid is closed. **(c)** A scanner unit with closed lid, as during operation, shown from rear to highlight fans mounted into the chassis. **(d)** The typical plate configuration consists of 16 plates inserted face down through holes punched into a rubber mat. The mat seals the plates to the glass surface of the scanner, stabilizing their position and preventing dessication. Note the light source located in the lid (trans-illumination mode).

S1.2 Temperature control during image acquisition

An important aspect of our method is the control of temperature within a tight range, reducing the large excursions that are sometimes unavoidable in the manual method where plates are moved in and out of an incubator for scoring. The notorious temperature sensitivity of *C. elegans* is highlighted by its variation between scanners (Fig. S1-2).

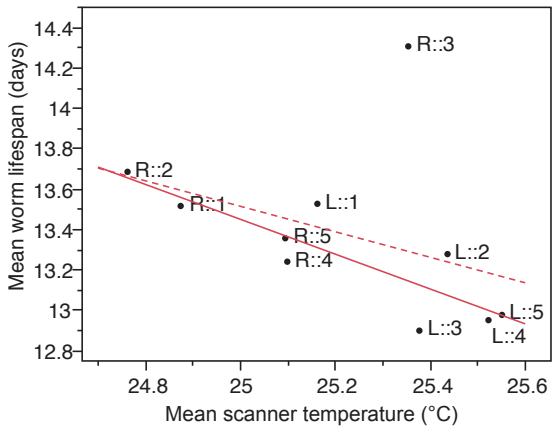

Figure S1-2: The effect of scanner temperature on lifespan. The lifespans of wildtype animals were acquired as described for Fig. 3 of the main text, and loaded onto 160 plates distributed across 10 scanners. The average lifespan of all animals on a given scanner was compared to the temperature measured on that scanner (Online Methods). A linear regression was run with all scanners ($r^2 = 0.163$, red dashed line) and with the R::3 outlier excluded ($r^2 = 0.746$, red solid line).

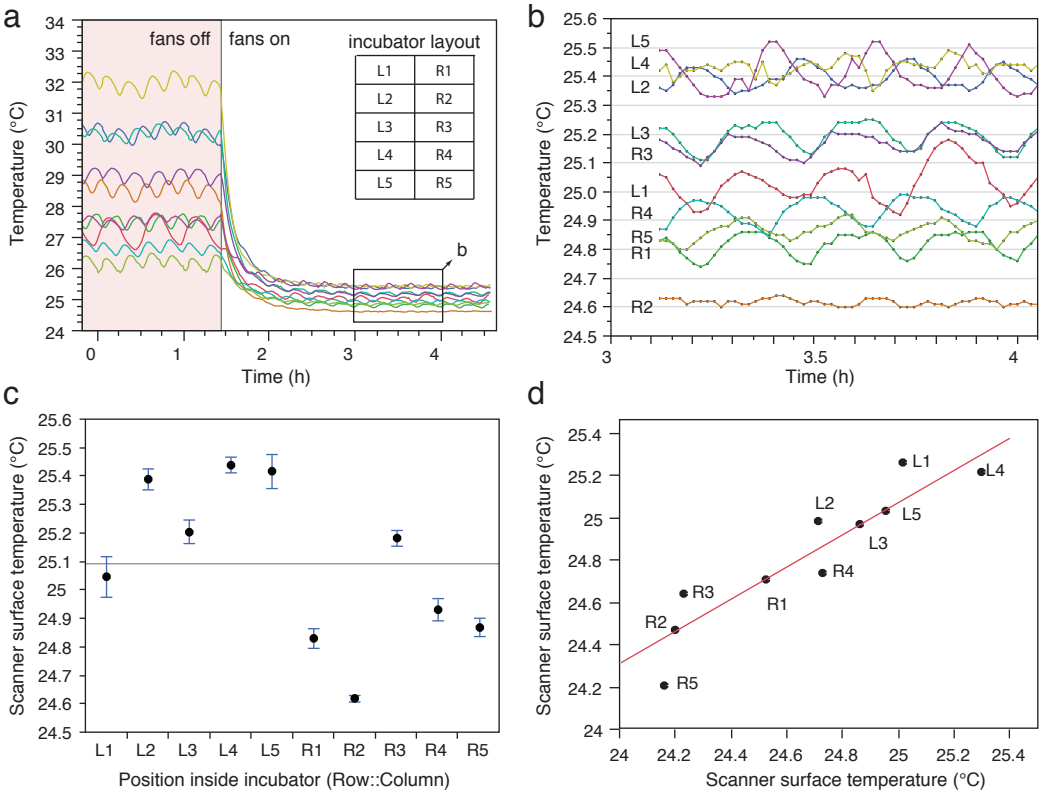

Figure S1-3: Consistency of scanner surface temperature. Caption on next page.

Figure S1-3 (Cont'd): **(a)** Local temperature differences between and temperature excursions of 10 scanners operating at a scan frequency of 15 minutes inside an incubator (Online Methods). On the left, the fans are turned off; on the right, they are turned on. The inset diagrams the shelving of scanners for reference in panels (b)-(d). **(b)** Detail of the rectangular area indicated in panel (a). Fans reduce local temperature differences due to incubator airflow to within $1°C$ and temperature excursions from scanning to within $0.2°C$. **(c)** Data shown in panel (b) are reported against the location of the scanners in an incubator adjusted to maintain $25°C$, with 5 shelves and 2 scanners (Left and Right) per shelf; see inset of panel (a). Time series are summarized by the mean with error bars indicating one standard deviation. The left side of the incubator has a slightly higher average temperature compared to the right side due to incubator airflow. **(d)** In a separate experiment, the scanner temperatures of a single incubator were measured at one time (x-axis) and again two months later (y-axis). The temperature measures correlate ($R^2 = 0.85$), indicating stability of operating conditions.

Temperature control is especially important when measuring lifespans outside the "room temperature" range, as with thermotolerance assays (e.g. Fig. 5 of the main text). To control temperature gradients that result from a scanner's electronics and light source, cooling fans are mounted into the chassis of the scanner, as shown in Fig. S1-1b and schematically in Fig. 1 of the main text. Fig. S1-3 below shows the efficacy of these modifications.

Supplementary Note S2 — Image processing

S2.1 Masking, background subtraction, and segmentation

In images captured by a scanner, only areas corresponding to the agar surface of each plate need to be analyzed. "Masking" is the process of separating multiple plates and removing their edges. The Worm Browser (Note S4 section S4.2 and Supplementary Video 6) generates a low-resolution composite of each image captured. The user can then draw on top of this image to produce an overlay (the mask) that specifies the location of each plate. This drawing can be done quickly in programs such as Photoshop or GIMP [Jones, 2008]. (This task is a candidate for future automation.) The image-processing server uses the resulting mask to extract individual plate images and stores them in a stereotyped folder hierarchy associated with each experiment.

Prior to analysis, all images $X^{(t)}$, $t = 1, \ldots, T$, undergo background subtraction. Each pixel $x_i^{(t)} \in X^{(t)}$ is reduced by the median value in a surrounding square of 25×25 pixels $198\,\mu m \times 198\,\mu m$. This median filtering removes variation at spatial scales larger than roughly twice the average width of an adult worm ($198\,\mu m$) suppressing background effects including shadows, smooth bacterial lawns, and uneven agar shading.

Median filtering sets the stage for image segmentation—the separation of worms from their agar and bacterial surface background—using a global threshold. Two global thresholds are applied to images, one restrictive threshold that identifies only dark objects such as worms, but tends to break worms into multiple objects, and a second more permissive threshold that tends to pick up extraneous features of the bacterial lawn. The two are compared to produce a final image segmentation. See also Fig. S2-5.

S2.2 Registration of consecutive scanner images

The panels in Fig. S2-1 show the magnitude of two types of jitter in the acquisition of images by our scanners (and presumably consumer electronics scanners in general). Displacements like

these must be taken into account when processing images leading up to the movement analysis (sections S5.1 and S5.2).

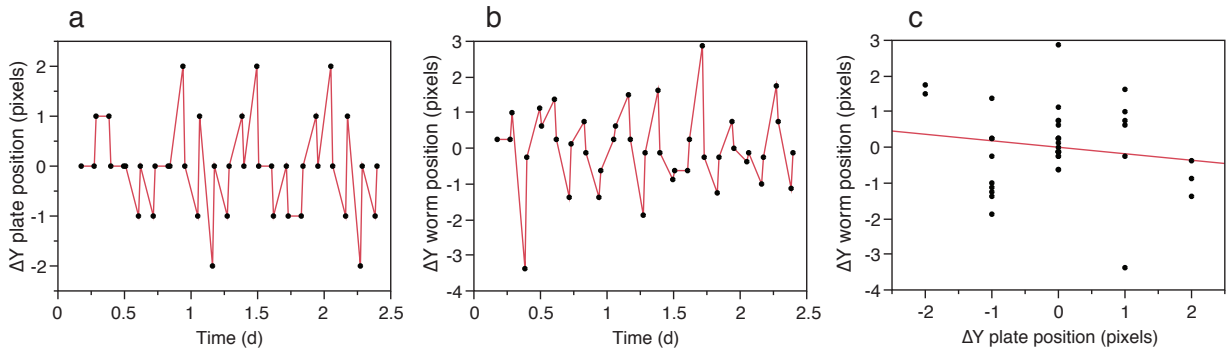

Figure S2-1: Registration of consecutive scanner images. **(a)**: The mechanism advancing the scanning bar causes offsets of a few pixels relative to world coordinates between subsequent images of the same plate. **(b)**: Slight distortions in plate images cause displacements of a few pixels relative to plate coordinates for stationary worms. **(c)**: The two types of displacements shown in panels (a) and (b) are uncorrelated.

S2.3 Worm recognition

The image-processing pipeline computes for each foreground object a set of features that are subsequently used by a Support Vector Machine (SVM) [Boser et al., 1992], Fig. S2-4, to discriminate between worm and non-worm objects, Fig. S2-3. The following is a list of 65 features used by SVM-based worm classifier. Their distribution of values is shown in Fig. S2-2.

- Pixel Area
- Rectangular Width
- Rectangular Height
- Rectangular Diagonal
- Spine Length
- Average Width
- Max Width
- Min Width
- Variance in Width
- Width at Center
- Width at Front
- Width at Rear
- Spine (SP) Length / Pixel Area
- SP Length / Rect Width
- SP Length / Rect Height
- SP Length / Rect Diagonal
- SP Length / Max Width
- SP Length / Average Width
- SP Length / Width at Front
- SP Length / Width at Rear
- Ratio of end widths
- Average Curvature
- Max Curvature
- Total Curvature
- Curvature X Intercepts
- Curvature Variance
- Distance between ends
- Absolute Intensity (AI)
- AI Variance
- AI Skew
- AI Roughness (Entropy)
- AI Roughness
- AI Maximum
- AI Average Along SP
- AI Average in Nbrhd
- AI darkest 20% of pixels
- AI darkest 20% area
- AI Dist from Nbrhd
- AI Variance Along SP
- AI of Region
- AI Normalized Average
- AI Normalized Max
- AI Normalized Avg Along SP
- Relative Intensity (RI)
- RI Variance
- RI Skew
- RI Roughness (Entropy)
- RI Roughness
- RI Max
- RI Average Along SP
- RI Average in Nbrhd
- RI darkest 20% of pixels
- RI darkest 20% area
- RI Dist from Nbrhd
- RI Variance Along Spine
- RI of Region
- RI Normalized Average
- RI Normalized Max
- RI Normalized Avg Along SP
- Edge Area
- Edge Area / Object Area
- Intensity Profile at Edge
- Intensity Profile at Center
- Intensity Profile Max
- Intensity Profile Variance

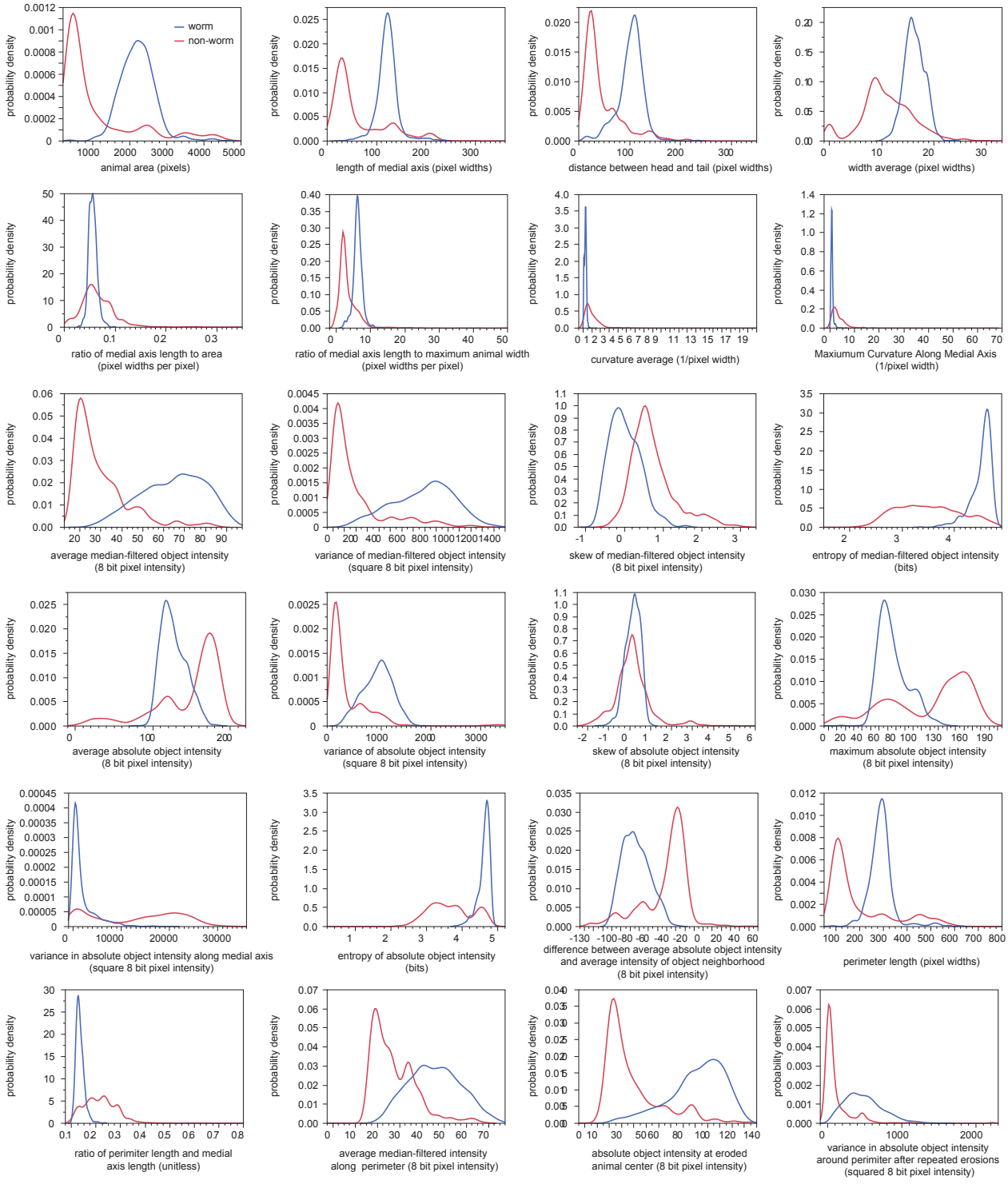

Figure S2-2: Feature palette for worm recognition. The panels show the frequency distribution of the values for each feature in the case of worms (blue) and non-worms (red). Because feature value distributions of worm and non-worm objects overlap in all cases, no single feature is sufficient for accurate discrimination. 1 pixel width corresponds to 8μm.

Because no single feature is sufficient to accurately discriminate between worm and non-worm objects, information from multiple features must be integrated. To assess whether accurate discrimination is feasible, we hand-annotated ~4,000 images as depicting either a worm or a non-worm object. We divided this set into two parts, a training set (1489 worms, 945 non-worms) and a test set (992 worms, 631 non-worms). Using the `libsvm` library, we built an SVM model of the training data, and evaluated its performance on the test set. The confusion matrix is shown below, demonstrating a discrimination accuracy of 97.7%. The Worm Browser (Note S4 section S4.2) allows rapid inspection of all objects predicted by the machine to be worms. Mistakes can therefore be annotated, producing a final false positive rate of nearly zero.

		SVM prediction	
		worm	**non-worm**
by-hand	**worm**	991	1
	non-worm	36	595

To visualize the feasibility of discrimination, we computed the eigenvectors of the feature correlation matrix of the gold standard and represented each animal as a dot projected onto the subspace spanned by the eigenvectors with the three largest eigenvalues (principal components PCA 1-3), as shown in Fig. S2-4.

Figure S2-3: Worm and non-worm objects. The images inside the red rectangle exemplify worm objects; the images on the left, outside the rectangle, illustrate various non-worm objects ranging from fibers, to dust, to lot numbers stamped on plates. The two classes of objects are well-separated by the Support Vector Machine, Fig. S2-4, based on the feature set shown in Fig. S2-2.

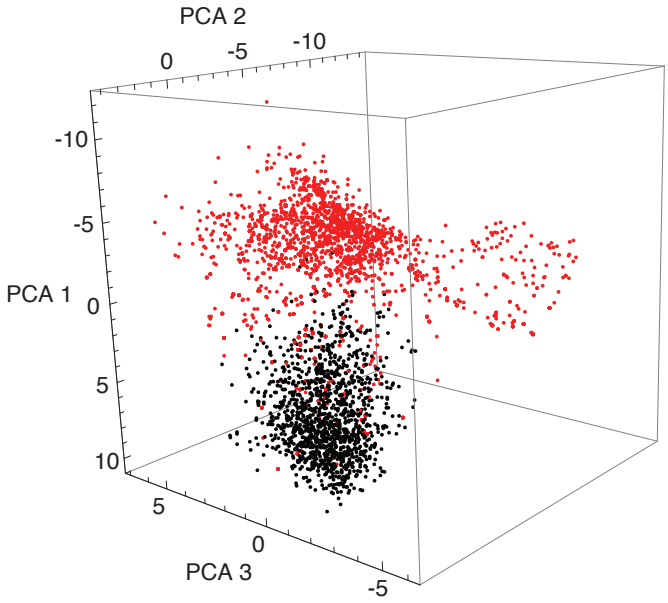

Figure S2-4: Support Vector Machine for worm recognition. Black: worms, Red: Non-worms. Each object of a hand-annotated "gold standard" is represented as a dot in a 65-dimensional feature space. The figure shows the projection of this point cloud onto the subspace spanned by the three principal components that explain the largest share of total variation (60.4%).

The previous table was based solely on wildtype. The following table illustrates the performance of the SVM when trained on a mix of strains and then applied to each separately (rather than trained on each specifically). For this analysis, we (1) lumped a number of genotypes into one big set; (2) assigned randomly 3/5 of animals to the training set and the remainder to the test set; (3) trained an SVM on the training set; (4) tested it on the test set; (5) calculated the confusion matrix for each genotype. FP=false positives, FN=false negatives:

	all	***daf-16 (mu86)***	**wildtype**	***glp-1***	***age-1(hx546)***
FP	84/1550 = 5.4%	23/207 = 11.1%	21/562 = 3.7%	32/453 = 7.1%	8/328 = 2.4%
FN	73/2426 = 3.0%	11/336 = 3.3%	19/777 = 2.4%	31/883 = 3.5%	12/430 = 2.8%

All strains have a false negative rate between 2.4% and 3.5%. The false positive rate is more varied, but ultimately immaterial, since we catch the non-worm objects during visual inspection in the quality control phase with the Worm Browser.

Fig. S2-5 provides examples of image segmentation and object recognition using the SVM, illustrating the difficulties of disentangling worm clusters. Refer to sections S8.2 and S8.3, as well as Figs. S8-3 and S8-5, for more detail on the accounting of multi-worm clusters. Supplementary Videos 1–5 provide time-lapse movies of plates with wildtype, *age-1(hx546)*, and *unc-64(e246)* mutants for the entire duration of an experiment, both with and without metadata overlay.

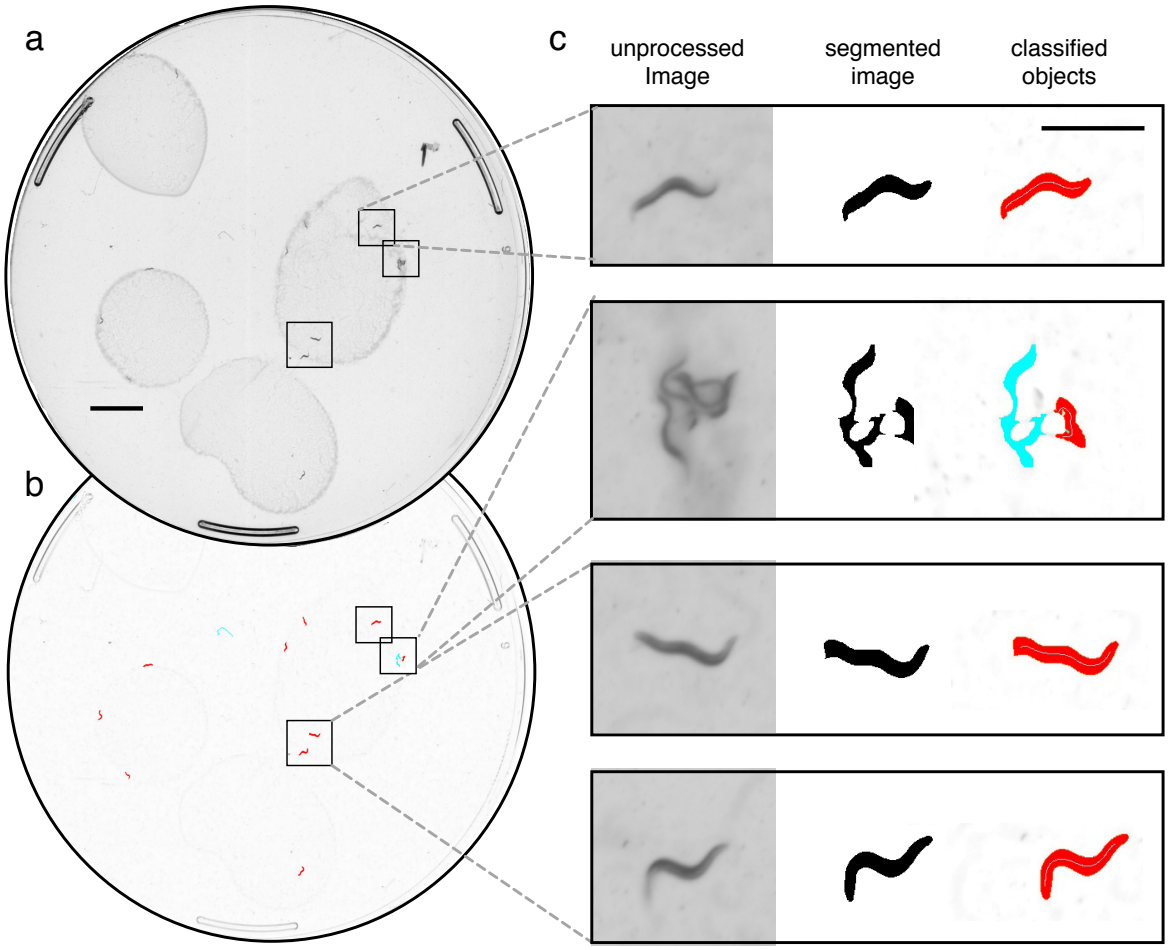

Figure S2-5: Image segmentation examples. Caption on next page.

Figure S2-5 (Cont'd): **(a)** A scanner raw image of an agar plate with worms on patches of bacterial lawn (*E. coli* OP50). Scale bar on the left corresponds to 5mm; the scale bar on the right to 1mm. **(b)** The same image after background subtraction (using a median filter), with foreground objects resulting from image segmentation highlighted in color. Objects classified as worms by the SVM are colored red; non-worms are colored blue. **(c)** Enlarged images of worms indicated in panel (b), as they appear in consecutive image processing steps. The scale bar marks 1mm. The first, third, and fourth rows are examples of correctly processed animals. The second row illustrates the difficulties of parsing multiple worm clusters. Note how several of the worms are incorrectly identified. Such errors, when they occur in the context of dead worms, can be rapidly identified in the Worm Browser (Supplementary Video 6).

Supplementary Note S3 — Assembly

S3.1 Introduction

The lifespan machine identifies worm death times through the combination of four interacting components:

(i) An image acquisition system based on modified flatbed document scanners;

(ii) An image acquisition control server running under Linux on a PC;

(iii) An image analysis server that can run in the background of any x86 Windows, Apple, or Linux computer;

(iv) A data annotation and aggregation client program that allows rapid inspection and annotation of automated results.

This document contains specifications for a reference implementation of the image acquisition system. An up-to-date, detailed specification of the software is provided as part of the source code documentation.

These instructions are lengthy, representing the authors' attempt at assembling a comprehensive, self-sufficient document. The modifications themselves are actually quite straightforward. A research assistant (RA) with a life science undergraduate degree and no engineering background learned how to perform all modifications over the course of several hour-long training sessions. The modifications required to complete two scanners, capable of monitoring over a thousand individuals, should take a novice two or three half days. All modifications are easily performed on a standard lab bench.

The authors do not currently possess good estimates for the operating lifespan of a modified scanner. Existing devices appear to operate correctly after nearly two years of continuous use. Several have been in operation for nearly five years. Across a fifty-scanner installation, the

authors have retired only three scanners as a result of malfunction, usually after only a few weeks of operation. Presumably, scanner fluorescent bulbs will eventually require replacement.

S3.2 Complete parts list

The following list details the equipment required for a reference, 10 scanner installation capable of monitoring 5600 nematodes simultaneously. Smaller or larger clusters can be easily assembled at proportional cost. All prices are from circa 2012 and approximate.

10x Epson v700 Perfection V700 Photo Scanners.

Available from Amazon.com or Epson.com for $540. In the past, we have also used the discontinued Epson 4990 Photo scanner. Other scanner models that possess appropriate transillumination capabilities are likely to be adaptable.

70x 80x80x25mm DC fans

This is a standardized computer chassis component, and as such is widely available from many suppliers. For example, the Thermaltake DuraMax 8 AF0058 80mm Case Fan available from NewEgg.com for $8.20.

10x 120x120x25mm DC fans

This is a standardized computer chassis component, and as such is widely available from many suppliers. For example, the Thermaltake A2492 120mm Case Fan is available from NewEgg.com for $9.

1x A 300+Watt source of 12V DC Power

We use standard Power Supply Units (PSU), which are mass-produced components of PC computer chasses, and as such are widely available at low prices from many suppliers. For example, the CORSAIR Builder Series CX430 V2 (CMPSU-430CXV2) 430W is available from NewEgg.com for $45

1x Temperature-controlled environment for equipment

This might be an incubator or a temperature-controlled room. For example, the Thermo Forma Environmental Chamber is a multi-use incubator which many labs already possess.

This incubator, in 2010, cost approximately $13,000 new, and as such represents more than half the price of an incubator-based Lifespan Machine installation.

50x Machine Screws

For example, Stainless Steel Machine Screw and Hex Nuts 11/32x1/8" 8-32 available from McMaster-Carr; a 100-unit box for $4.25

50x Machine Screw Nuts

For example, Stainless Steel Machine Screw and Hex Nuts 11/32x1/8" 8-32 available from McMaster-Carr; a 100-unit box for $4.25

80x Nylon twine or string, cut into 16.5" lengths

Used for mounting fans. Simpson's Individual Stringettes; available from McMaster-Carr; 1800 feet for $8.00

1x Reasonably modern (circa 2012) x86 PC or laptop capable of running Linux

More details provided under "Software Components" (section 1.3.1 of the scientific Supplementary Information); $650

10x 9x11.5" inch glass sheets

Standard window pane glass, available from most local hardware stores; $5 per sheet.

10x 8.7x11.3" inch rubber mats

Cut from rubber stock, for example High-Strength EPDM Rubber Sheet, 40A,Black, 36"x36"x1/16" available from McMaster-Carr at $20 per foot.

40x Plastic spacers to support the scanner lid

1" segments cut from 8' Polyethylene U-Shaped Channels. Leg Length - 3/4", Base Length - 5/4". Available for McMaster-Carr; $18.8 for eight feet.

2x Gaffer's tape

This is similar to duct tape, but better in every way. Available from McMaster-Carr; $10 a roll. Black looks sleek; pink looks totally rad.

10x Wire to power fans

20 AWG gauge or lower.

10x Rubber mats

Plates should be sealed to prevent desiccation. Many methods for doing this are possible, we use a design cut into High-Strength EPDM Rubber Sheet, 40A,Black, available from McMaster-Carr at $22 per 3 feet. A reference design, appropriate for laser-cutting, is provided.

1x Round Hole Arch Punch 2" Diameter —*Optional tool for construction*

If a laser-cut design is not available, rubber mats can be manufactured by hand using this punch. Available from McMaster-Carr for $70.

160x Small Petri Dishes —*Consumed by each 5600 worm experiment*

BD Falcon Petri Dishes, 50x9mm, Tight-Fit Lid, Sterile, BD Biosciences. Available from multiple suppliers; $215 per box of 500.

40x Butt-splices or wire nuts —*Optional*

To avoid soldering joints, butt-splices or wire nuts can be used to splice fan wiring. Available from McMaster-Carr in large packs for a few dollars. Yes, these have a silly name.

1x Mini screw driver set for scanner disassembly —*Tool for Construction*

Already present in most labs. For example, a 5-Piece Phillips Screwdriver Set available from McMaster-Carr for $22.18.

1x Spiral saw for cutting holes in scanner chassis —*Tool for construction*

Other types of cutting devices may work well, but a spiral saw cuts quickly and will impress your friends. For example, Rotozip Spiral Saw Electric, 5.5 Amp, 15000 - 30000 rpm, 9-1/4" L, available from McMaster-Carr for $106.25.

1x Hot-Melt Glue Gun —*Tool for construction*

Standard Duty Hot-Melt Glue Gun available from McMaster-Carr for $23. Packs of Hot melt glue sticks available from McMaster-Carr for $10.

1x Soldering Iron or Gun —*Optional tool for construction*

For example, a Soldering Gun available from McMaster-Carr for $58

1x A power drill / screw driver —*Optional tool for construction*

Unless you have wrists of steel, it is easier to use a power drill to add and remove machine screws. Available from McMaster-Carr for $150, or for less at a home improvement store like Lowes.

1x Canned air —*Optional tool for construction*

A source of compressed air, either from a can (often sold to clean computer keyboards, sold at any office supply store), or from a lab air outlet.

10x Additional 80x 80x 25mm DC fans —*Optional*

Used to circulate air throughout incubator to achieve best-possible temperature calibration. Only required for certain experimental designs.

10x Metal ductwork —*Optional*

Used to circulate air throughout incubator to achieve best-possible temperature calibration. Only required for certain experimental designs. For example, 3" Aluminum Bend-and-Stay Duct Hose, available from McMaster-Carr for $5.25 Each.

2x Plastic bin for precise drying of plates —*Useful ancillary equipment*

Exact dimensions are not crucial, but long and flat boxes work best. For example, the 35-5/8" x 18-1/4" x 6-1/4" h Long Underbed Box with Wheels, available from The Constainer Store . com for $20.

2x Sheet of Wire Mesh, fitted to plastic bin (i.e. 18"x36") —*Useful ancillary equipment*

Mounted in plastic bins for precise drying of plates. Available at local hardwarde stores. Corrosion-Resistant Mesh Fencing is available from McMaster-Carr for $4.70 per square foot.

1x 25lbs Regular (non-indicating) 8-mesh Drierite —*Useful ancillary equipment*

Available from drierite.com for $83.38.

S3.3 Recommended scanner modifications

In this section we present the steps required for modifying stock Epson v700 Perfection Photo Scanners into a device capable of producing data of the quality presented in our manuscript. The

modifications accomplish two goals: (1) install fans as a means for regulating scanner surface temperatures and (2) raise the focal plane of the scanner to optimize its use in scanning agar plates.

We see great potential for the future development of simpler, more effective strategies to accomplish these goals. The authors consider the current design primarily as a starting point from which the technology can be adapted according to each individual researcher's needs.

Overview of the fan mounting procedure

The Perfection v700 Photo Scanner has two parts—a bottom chassis, which contains the CCD image sensor, and a detachable lid (also called the Transparency Unit or "TPU"), which contains the transilluminating light source. To prevent the build-up of heat within the chassis, we mount three 8cm computer fans into the sides of the bottom chassis (labelled F1 ,F2, and F3 in the figures). These serve to move the heat produced during each scan out of the scanner body. To prevent the build-up of heat within the TPU, we hang four fans (F4, F5, F6, and F7), which serve both to remove heat from inside the TPU and to circulate cold air across the plates. A single 12cm fan (F8) is hung from the rear of the TPU, which serves to remove heat from the scanner TPU, the scanner chassis, and moves cold air across the plates.

Four cuts are made into the scanner chassis, and three into the TPU, to permit cold air to cycle through both compartments. These cuts are easily made using a spiral saw or other cutting device. Our experience is that most Dremel tools aren't quite up to the task of cutting through the hard chassis plastic. When working with any cutting tool, wear safety goggles and always pay attention. Clean, uncluttered work spaces promote safe practices.

Three fans—F1, F2, F3—are sealed in place using a glue gun. Four fans are hung from the scanner TPU are suspended by a loop of nylon twine, hanging from machine screws drilled into the TPU at positions S1, S2, S3 and S4. The holes can be cut with a power drill if one is available, or else using the spiral saw.

The exact location and dimension of each modification are presented in Figure S3-1. Photographs of the cut holes and glued fans are provided in several consecutive figures.

Modifying the scanner chassis

(1) Connect an assembled scanner to a computer and scan a piece of paper, to confirm that the hardware functions. Scan in "transparency" mode to confirm that the light source in the TPU functions properly.

(2) Disconnect the scanner and remove the TPU.

(3) The top surface of the scanner chassis is secured by several screws, located under small plastic covers in each corner. Pop out the plastic caps and release the screws. Pop off the top surface of the scanner and place it somewhere safe. Be careful not to get finger prints on the bottom of the scanner glass.

(4) Using your hand, gently push the scanner bar 3/4 of the way toward the front of the scanner, to move it out of harm's way.

(5) Place a piece of tape over the long slit opening of the scanner bar, to prevent plastic chips from entering the bar.

(6) Using a spiral saw or your favorite cutting device, cut holes H1, H2, H3 and H4. Cuts H1 and H3 run somewhat close to internal scanner wiring. The cuts shouldn't come extremely close, but be careful and fold any nearby cables out of harm's way. Any time you use a spiral saw, wear safety goggles and stay aware. The spiral saw tends to throw little chips of plastic.

(7) Take an 8cm fan, and use it to confirm that a fan will fit snug into each hole. Enlarge the holes where necessary. However, don't remove more plastic than required, as doing so will make it harder to hot-glue the fans in place.

(8) Using compressed air, either from a can or from a hose hooked up to house air, blow away all the plastic chips produced by the cuts. If ignored, these plastic chips have the nasty habit of working their way into moving parts, in particular the scanner's power button. If this button ever stops working, there's probably a chip in it somewhere.

(9) Place a 8cm fan into hole H1. Make sure the power cable is facing towards the front of the scanner, preferably from the upper right hand corner of the fan. Fan F1 should face outwards—the airflow should be heading into of the scanner. Glue it in place along all four sides. You may need to hold the fan in place with one hand and glue with another. In the same way, glue fans F2 and F3 into holes H2 and H3. Fan F2 and F3 should both be facing

outwards.

(10) Remove the protective tape added earlier to the scanner bar.

Modifying the transparency unit

(1) Put the scanner chassis aside, and consider the scanner TPU. The bottom of the TPU is secured to its the cop cover by eight screws. Release them and pop off the bottom cover, and place it somewhere safe

(2) Make cuts H5–H7. Note that the TPU connector is located at the center of H7. Remove this cable from its notch, and make cut H7. You might consider cutting a new notch somewhere next to cut H7, so that this cable can be again secured.

(3) Using a power drill or the spiral saw, drill holes S1 through S5. Make sure to use a bit appropriate for the machine screws you are using, as you'll want the screws to fit snugly.

(4) Using compressed air, either from a can or from a hose hooked up to house air, blow away all the plastic chips produced by the cuts.

(5) Insert machine screws into each hole S1–S5. The head of the screw should be inside the TPU. Secure the screw by tightening a nut on the reverse side (the outside top of the TPU).

(6) Grab the scanner TPU bottom, and snap it back into place. Screw in all screws.

(7) Cut five 16.5" inch segments of nylon twine. Put the TPU of a scanner on a flat surface, with the machine screws facing upward. Thread one end of a piece of thread through the two holes at the top of a 8cm fan. Tie a knot at the appropriate piece of the thread such that the fan hangs snug at the TPU's side when hung from machine screw S1. Repeat three more times, hanging fans from machine screws S2–S4. Remember to orient the fans correctly; Fans F4 and F5 face inward; fans F6 and F7 face outward.

(8) In the same way, tie a 12cm fan F8 to machine screw S5. It should face inward.

(9) Finished!

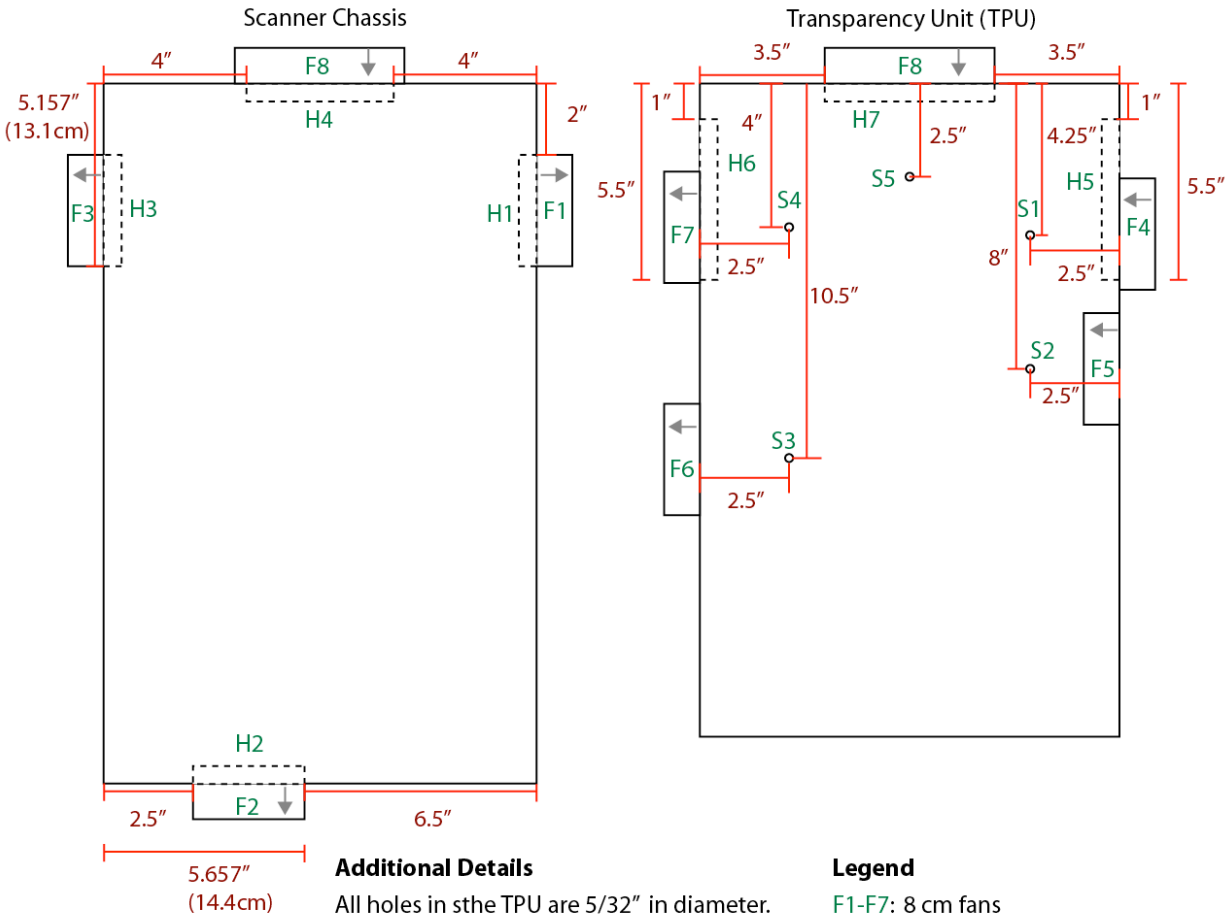

Figure S3-1: Scanner modification diagram. The location and dimensions for each cut are illustrated, for both the scanner chassis and the Transparency Unit (TPU).

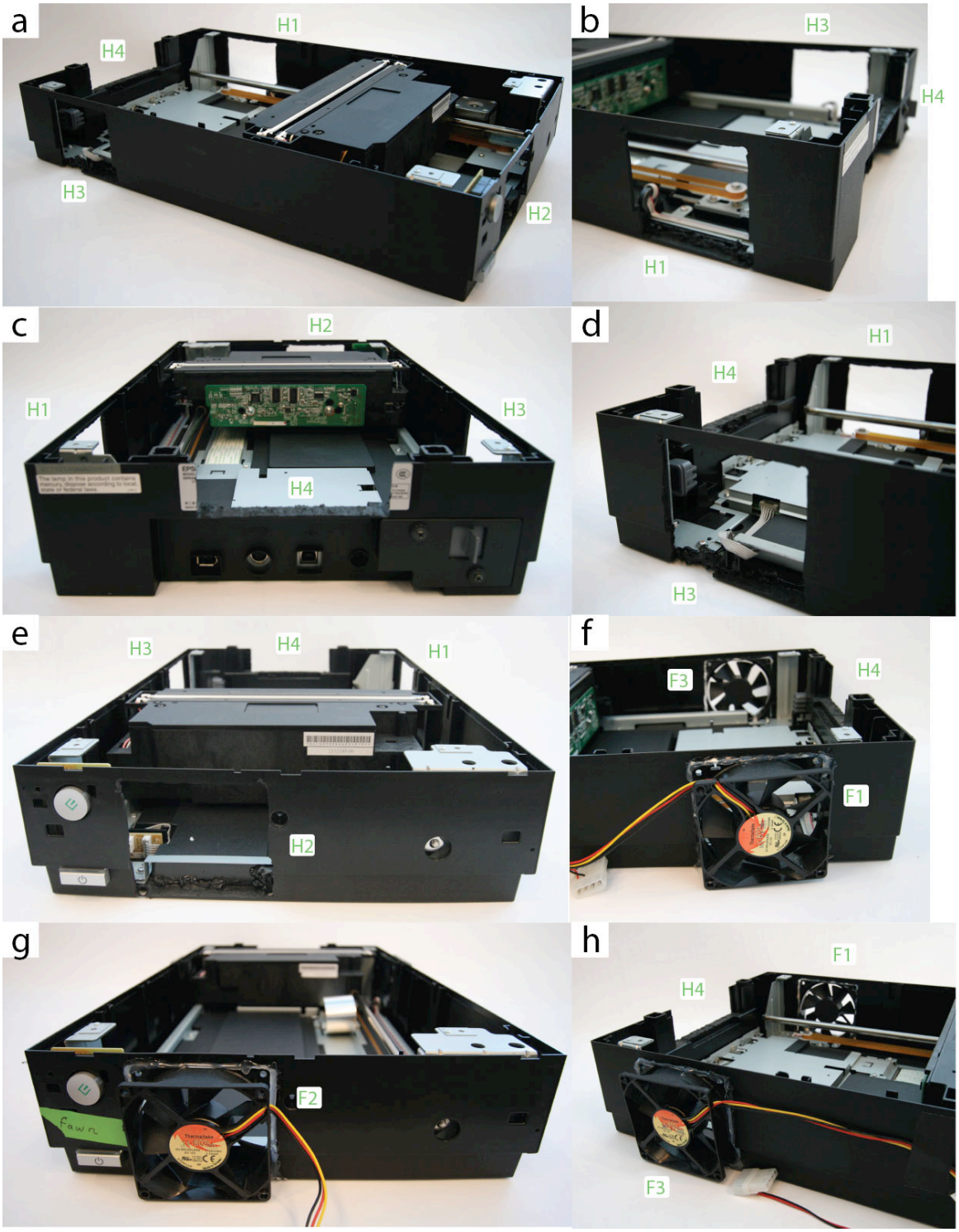

Figure S3-2: Photographs of chassis modifications. **(a):** The scanner chassis with its lid removed, and four cuts, H1–H4, made into its walls. **(b–e):** Close up views of each cut. **(f–h):** 8cm fans are glued into holes H1–H3.

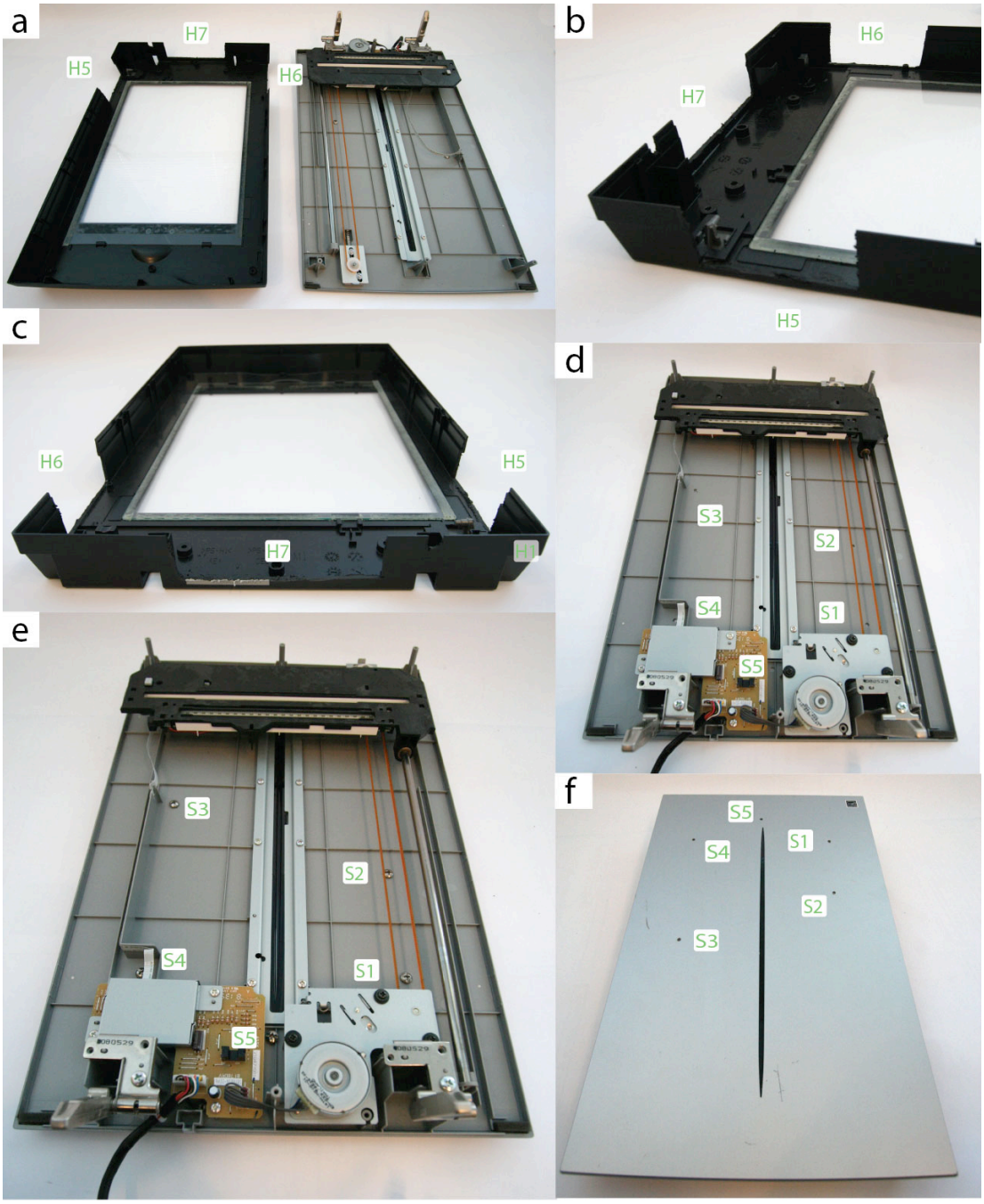

Figure S3-3: Photographs of mounted scanner fan. **(a):** The Transparency Unit (TPU) is shown with its top and bottom separated. **(b–c):** Three cuts, H5–H7, are made into the bottom of the TPU. **(d):** Four holes are drilled into the TPU lid, and **(e)** filled with machine screws. **(f):** The holes are shown from the top of the lid.

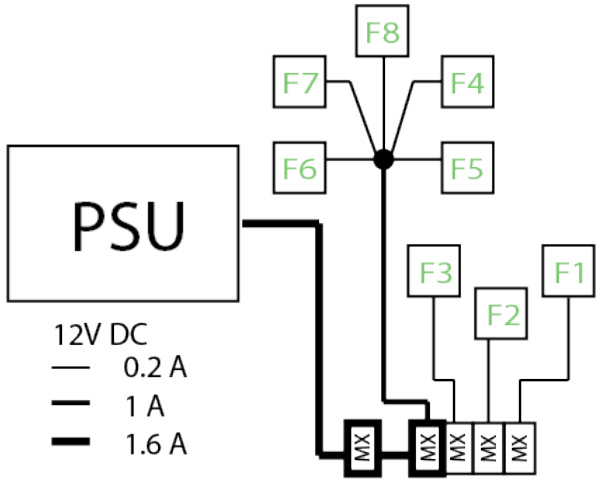

Figure S3-4: Fan wiring diagram. The wiring that supplies the fans is shown. "MX" indicates a standard Molex connector. The connection between fans F4–F8 can be made by clipping off the Molex connectors and splicing the wires together using solder, or by using a butt-clamp or a wire nut. Be careful to add appropriate insulation. Be careful to plug the Molex connectors in the correct order, otherwise excess current may be passed through each connector. Be careful to use an appropriate of gauge wire for the indicated current. The eight fans attached to each scanner draw about 1.6 amps at 12V DC. To prevent heat build-up along the wire, a voltage drop of less than 3% along the wire is widely considered safe. If the power source is 10 feet away from the scanner, this suggests that an AWG gauge no larger than 20 should be used. Incorrect wiring can generate excess heat and represents a fire hazard.

Accessing and preparing the fixed scanner lens

The Epson Perfection v700 contains a single fixed lens, whose position sets the focal plane of the device. Worms on agar plates are out of focus on stock scanners, because they are raised above the standard focal plane. To optimize image quality, we move the lens position very slightly.

(1) The scanner lens is located inside the moving scanner bar, situated inside the bottom scanner chassis. Remove the lid of the scanner chassis (as described in Recommended Mounting Procedure Step 3), and identify the scanner bar. Refer to Figure S3-5.

(2) There is a removable plastic flap in the top of the bar. Remove it with a screwdriver. Put it aside in a safe place. If you look into this slot, you should see, a few inches down, a black tube. This is the top of the lens. Unfortunately, scanners ship with this lens glued in place by a piece of hard resin. We need to remove this resin before we can adjusting the lens

position. Removing the resin, in turn, requires we access the lens directly through a port on the bottom of the scanner bar. This, in turn, requires we disconnect the scanner bar from the scanner chassis, to make it possible to access the bottom port.

(3) Identify the small metal clip that attaches the scanner bar to the drive belt. The drive belt is the long rubber belt running the length of the scanner chassis. Take a look at this clip and remember how it is attached. Pop it off by pushing down on the exposed lever.

(4) Locate the nut at the front of the scanner that holds the support bar in place. The support bar is the long metal pole that running the length of the scanner chassis. Remove the nut. Notice that the bar is still held securely by a friction connection at the back of the scanner.

(5) When handling the support bar, wear gloves to avoid exposing it to skin oils. Rotate the support bar in place to release the friction connection. Sometimes, pliers may be necessary. Only use pliers on the rectangular end of the bar, never on the cylindrical part. The scanner bar and support bar should now be loose

(6) Remove the support bar from the scanner bar, and put it in a safe, clean place.

(7) Flip the scanner bar so that rests upside down inside the chassis. Be careful not to catch any wires or to crease the ribbon cable.

(8) You should see several screws on the bottom of the scanner bar, securing the bottom lid. Remove them, and remove the black bottom lid.

(9) You should now be able to see the fixed lens, as shown in Figure S3-5. It is black, often with a white dot. There is a second lens, used for IR measurements, which you can safely ignore. You should see a little piece of gum melted onto the lens. Remove it with a screwdriver or a razor blade. This should take about a minute. You'll need to work around the edges, but usually chunks can be pried off in big pieces. Make sure to clean around the edges of the lens.

(10) You have freed the lens! Replace scanner bar's bottoms cover, and secure all the screws. Make sure the ribbon cable running into the bar enters through the correct slot.

(11) Slide the support bar back into the scanner bar, and fit everything back into its original place.

(12) Rotate the support bar in situ such the friction joint holds it secure.

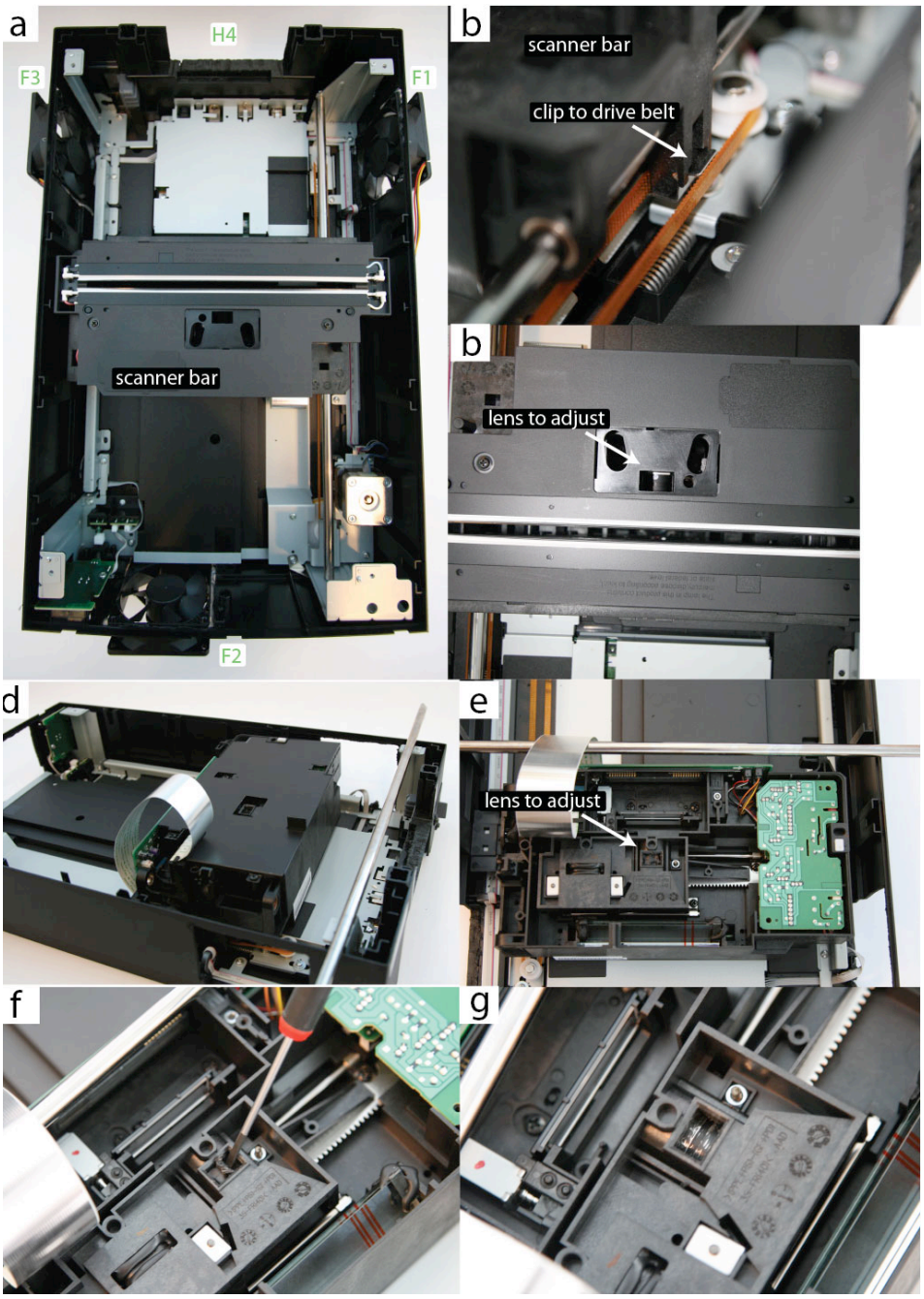

Figure S3-5: Inside the scanner chassis. **(a):** A top-down view of the scanner chassis with its lid removed. **(b):** A close-up of the clip that connects the scanner bar to the drive belt. **(c):** The slot in the scanner bar through which the lens can be adjusted. **(d):** The freed scanner bar can be overturned to reveal a bottom access panel. **(e):** With the panel removed, the bottom of the lens can be seen, glued in place. **(f):** The gum can be removed with a screw driver. **(g):** The lens freed from gum.

S3.4 Instructions for focusing a scanner

(1) Obtain a few reference plates with which to calibrate the focal plane. The authors' prefer to optimize focus using live worms on agar plates, but can imagine a more stable reference might be ideal. If live worms are used, let the plates dry enough so that they will absorb some liquid. Seed with 125–250 microliters bacteria and let the liquid absorb, providing the animals with a comfortable, dry, lawn.

(2) Pick 20–40 animals onto each plate.

(3) Very Important! Let the animals calm down for an hour or two at room temperature after plating, to ensure that they have stopped moving quickly. If calibration is attempted using startled worms, all images will appear blurry regardless of focus, due to animal motion. It is easy to waste an hour attempting, in vain, to focus fast-moving worms.

(4) If the scanner chassis lid is currently removed, place it back on. Do not tighten any screws, because it will be necessary to enter the chassis in each step of the focusing process. Do make sure that the lid is seated flush. Sometimes, the lid can catch on one side, which will raise it slightly and affect the focus.

(5) If the TPU is currently unattached, put it on and plug it in.

(6) If the scanner isn't plugged in to a computer and a power outlet, do so.

(7) If the chassis and TPU fans aren't running, plug them in and start them running. Scanning produces heat, and without operating fans, plates will begin to fog over during calibration.

(8) Place the calibration reference plates into the scanner. If plates are to be placed on glass sheets during experiments (as is recommended), remember to include this glass sheet into the calibration process.

(9) On the computer, run a preview scan that captures the entire scanner surface.

(10) Select a small location of one plate on which to focus. Smaller areas will be captured more rapidly, speeding the calibration process. If worms are being used, try to select a small area that contains several worms. The calibration process is fastest considering a single plate, but this is somewhat risky as plates may vary slightly in height. Ideally, the focal plane should be adjusted to a consensus height that works well for most plates.

(11) On the computer, capture an image.

(12) Inspect the captured image at 100% zoom. The quality of the current scanner focus can be evaluated through comparison to a collection of gold-standard images, available in Figure S3-6.

(13) If optimal focus has not been achieved, remove the lid of the scanner chassis, along with the entire TPU, as one piece. With proper organization of your workspace, it should be possible to do this without detaching any cables.

(14) Look through the top access port of the scanner bar, and identify the fixed lens. With a small flat head screw driver, nudge the lens slightly in the desired direction. The lens should be adjusted in tiny, nearly imperceptible increments. If you think you've made a large adjustment, in all likelihood the scanner is now extremely out of focus. In this case, it will be a long job hunting your way back. One strategy is to scratch notches into the paint of fixed lens housing. These scratches can then be used as a reference to mark your position.

(15) Reattach the chassis lid and TPU, making sure that the lid is seated flush.

(16) Repeat steps 11–15 until a scanner is in focus. It is the authors' experience for half the scanners they have adjusted, optimal focus is achieved within two or three adjustments. In this case, focus is achieved in less than five minutes. For the other half of scanners, optimal focus is more elusive, requiring up to half an hour of iterations. Be patient, this process only needs to be done once in the lifespan of the scanner.

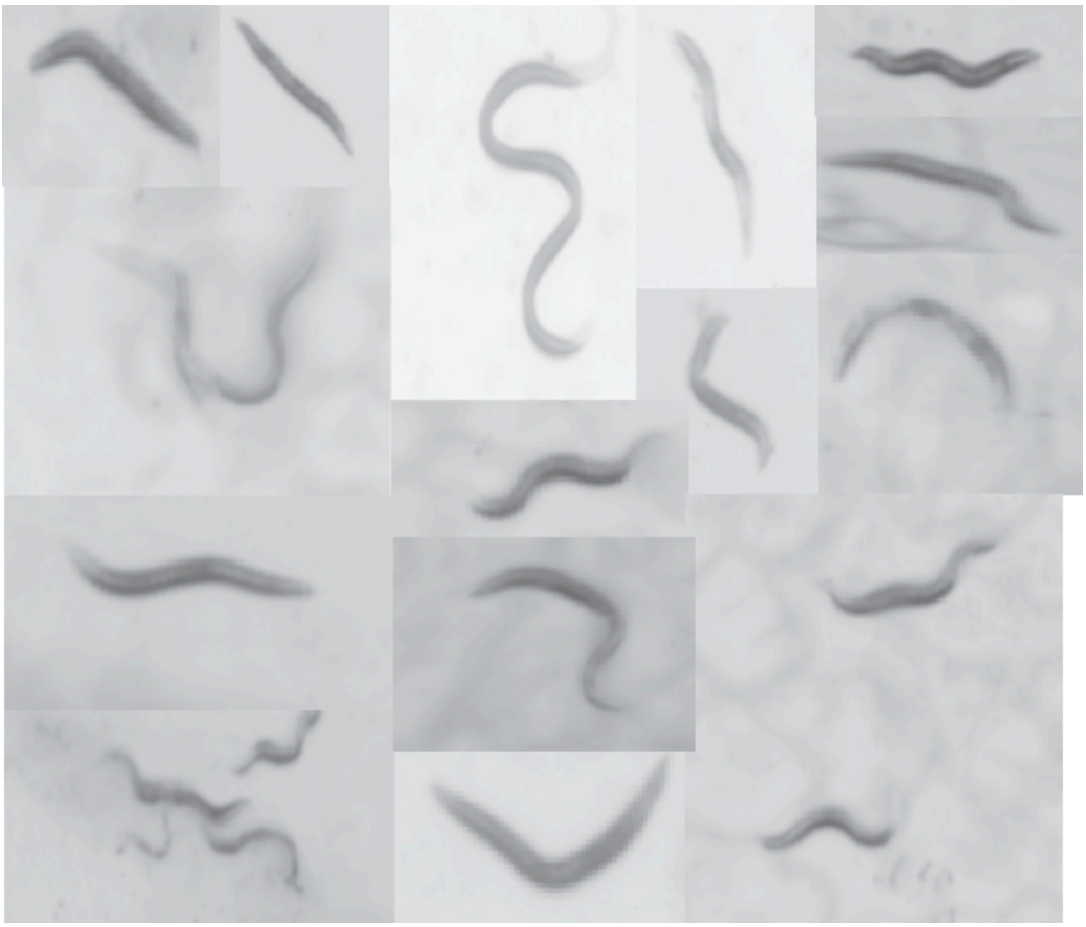

Figure S3-6: Reference images demonstrating sufficient focus. These images were collected across several scanners, and represent the end-point of scanner focal-plane optimization for the devices. In the authors' experience, attempts at acquiring images sharper than those presented are generally counterproductive.

Supplementary Note S4 — Software toolset

S4.1 Software components

1. *Statistical software:* To assemble survival curves, we used JMP [Sall et al., 2012], a commercially available statistical package, and R [R Development Core Team, 2009], an excellent GPL-licensed statistical software environment. In R, we used the surv, rms, flexsurv, and muhaz packages.

2. *Worm Browser:* See section S4.2.

3. *Image acquisition server software:* Custom software controls experiments by communicating with the scanners via USB. Scanning schedules are stored in the MySQL database and take roughly the form "At time t, run a scan on device D between positions (x_1, y_1) and (x_2, y_2)". Typically, four plates are imaged in one scan, representing a column of plates on the scanner. The server monitors the schedule database and executes scans as they become due. Various telemetry is collected, including the start and stop times of each scan, various gross image properties such as average image intensity, and a description of delays or errors that occurred. Since a Lifespan Machine may run for three weeks or more, it is likely that errors will crop up in some system component, such as a scanner, the network connection, a power supply, etc. The image server software has several layers of redundancy and error handling to mitigate such problems, including the ability to send emails or text messages reporting new problems.

4. *Image-processing software:* The image processing pipeline and associated job scheduling software is written in C++. The Worm Browser is written in C++ using the library FLTK (GNU GPL V2) to provide cross-platform GUI functionality.

5. *Availability:* All software is currently (2013) accessible at www.lifespanmachine.org.

S4.2 The Worm Browser GUI client

The Worm Browser is designed as a desktop client that complements the web interface to the lifespan machine. The Worm Browser performs five separate tasks:

- *Processing and uploading image capture schedules.* The Worm Browser accepts and processes a high-level description of an experimental schedule, written as an XML file. This schedule file contains a description of the image acquisitions that should be performed, in the style of "scan all plates on 10 scanners, executing a scan every 15 minutes. Do this for 28 days". The worm browser compiles this file into a list of specific scan times, which it uploads to the central database.

- *Generating and processing mask files.* Individual plates must be identified in images of scans. The Worm Browser creates the mask files and saves them to local storage for annotation. The annotated mask file is then loaded back into the Worm Browser and the results submitted to the central database.

- *Inspecting and annotating worm image data.* The Worm Browser enables a user to browse through various representations of processed image data, including "storyboards" that contain images and videos of each identified animal as it goes through the death transition (Supplementary Video 6).

- *Output of statistical files.* The Worm Browser can produce a variety of statistical data germane to both the operation and the results of analysis.

- *Debugging image processing pipeline steps.* The Worm Browser acts as a platform for experimenting with new image processing techniques. This is often easier in a client than working with the image processing server directly.

Figure S4-1: Worm Browser screen shot. See also Supplementary Video 6.

Supplementary Note S5 — Movement analysis

S5.1 Identification of stationary worms

Once foreground objects have been classified into worms and non-worms (Figs. S2-2 and S2-5), the task is to identify when individual worm-objects have permanently stopped moving. The Lifespan Machine distinguishes two kinds of motion: (i) spontaneous locomotion across the agar surface and (ii) subtle posture changes that do not involve whole body displacement (also referred to as *local* motion). Accordingly, the identification of death events occurs in two phases. First, we identify worms that have stopped locomotion. Such worms are termed stationary. Second, images of stationary worms are further analyzed to identify cessation of spontaneous posture changes. The retrospectively identified cessation of all motion is then recorded as a death event.

A trained Support Vector Machine (SVM), see Fig. S2-2 and Fig. 2f of the main text, identifies the position (x, y) of worm objects in each image $X^{(t)}$. Let

$$P^{(t)} = \{(x, y) \,|\, \texttt{object at position } (x, y) \texttt{ in } X^{(t)} \texttt{ is an SVM-identified worm}\}$$

denote the set of these positions. Early in a lifespan experiment, when most worms move around rapidly, the scanning rate of two images per two hour period does not allow determination of which worm in $P^{(t)}$ corresponds to which worm in $P^{(t+1)}$. To identify death events, however, we do not require such early-stage detail. It suffices that worms become trivially trackable once they have ceased locomotion. These worms are then monitored for subtle head and tail displacements, as described in Note S5 section S5.2. Identifying the onset of a worm's stationary phase is a retrospective analysis. It results in a straight "time-track" (Fig. 2f of the main text) with an associated series of images that is then subject to posture analysis. While this is straightforward in principle, the intermittent locomotion of worms, their spatial interactions, and the noisiness of image-capture (Fig. S2-1) and worm recognition create challenges.

We construct time-tracks of stationary worms in a stepwise fashion and clean them up prior to

posture analysis. The time-track of a stationary worm w_i is defined as a sequence of locations:

$$w_i = [p_1 p_2 \cdots p_n]_i \text{ with } \; p_1 \in P^{(t_1)}, p_2 \in P^{(t_2)}, \; \ldots \; , p_n \in P^{(t_n)}, \;\; t_1 < t_2 < \cdots < t_n \qquad (7.1)$$

The time-ordered series of stationary positions $p_k \in P^{(t_k)}$ are meant to correspond to the same spatial location across images $X^{(t)}$. The time points need not necessarily be consecutive scans, i.e. the track may contain gaps. Such gaps can arise for several reasons, such as the temporary misidentification of the object at that position (a second worm might transiently cluster or cross, creating a structure not recognized as a worm) or the behavior of the tracking algorithm (below). These time-tracks are built up sequentially proceeding backwards from the last scan in the experiment toward earlier scans. Stationary tracks of worms that died early may not remain identifiable throughout the experiment, since a dead animal may dissolve over time and become unrecognizable by the SVM. Thus, despite proceeding backwards from a state in which all worms are dead, we might nonetheless have to initiate new stationary tracks as we encounter their ending.

Outlining the procedure requires a bit more notation. Since a worm's time-track is built over time, we decorate w_i, defined above, with a scan time t, i.e.$w_i^{(t)} = [p_1 p_2 \cdots p_n]_i$, to indicate the most up-to-date time-track as of time $t \leq t_1$. (Remember we proceed backward in time; at the next addition of a time point, the current p_1 becomes p_2 and the current t_1 becomes t_2.) To denote the *set*, rather than the *sequence*, of positions belonging to worm i, we write $\text{Set}(w_i^{(t)}) \equiv \{p_1, p_2, \cdots, p_n\}_i$. This allows us to express that no location can belong to more than one time-track: for all $i \neq j$, $\text{Set}(w_i^{(t)}) \cap \text{Set}(w_j^{(t)}) = \emptyset$. We refer to the most recently added component (youngest animal age) of the time-track of worm i as $w_{i1}^{(t)}$, that is, $w_{i1}^{(t)} = p_1$ of $w_i^{(t)}$. Throughout the process we maintain a set of time-tracks $W^{(t)} = \{w_i^{(t)}\}$, each track tentatively belonging to a single worm. Corresponding to $W^{(t)}$, there is a set of anchor points $W_1^{(t)} = \{w_{i1}^{(t)}\}$ containing the most recently identified points of each track. Finally, we maintain a set of positions that have not been assigned to a time-track retrospectively up to time t, $F^{(t)} = \{p \mid p \in P^{(\tau)}, \; \tau \geq t \text{ and } \; \forall i \;\; p \notin \; \text{Set}(w_i^{(t)})\}$.

Given sets $W^{(t)}$ and $F^{(t)}$, we need to identify which positions $P^{(t-1)}$, derived from scan $t-1$,

extend (backwards in time) existing time-tracks in $W^{(t)}$, or initiate a new time-track. This is accomplished by finding an assignment $f:\ W_1^{(t)} \cup F^{(t)} \mapsto P^{(t-1)}$ that minimizes the total Euclidean distance $||\cdot||$ between positions p_k in $W_1^{(t)} \cup F^{(t)}$ and their images under f:

$$\sum_k ||p_k - f(p_k)|| \stackrel{!}{=} \text{minimum} \tag{7.2}$$

where as many elements as possible must be assigned. If either $W_1^{(t)} \cup F^{(t)}$ has fewer elements than $P^{(t-1)}$ or vice versa, some positions must be left unmatched (i.e. they are matched to nil). This combinatorial optimization problem is of a well-known form that can be solved [Kuhn, 1955].

The new set $F^{(t-1)}$ (recall that we proceed from later towards earlier times) is formed from all elements of $F^{(t)}$ that were not matched to a $p \in P^{(t-1)}$ plus all $p \in P^{(t-1)}$ that have no match in $W_1^{(t)} \cup F^{(t)}$. Each element in $F^{(t)}$ paired with an element in $P^{(t-1)}$ initiates a new time-track (and thus identifies a new worm) that is added to $W^{(t)}$ in producing $W^{(t-1)}$. A $p \in P^{(t-1)}$ that pairs up with $w_{i1}^{(t)}$ extends the time-track of worm $w_i^{(t)}$, i.e. $w_i^{(t-1)} = p w_i^{(t)}$. In effect, p becomes the new $w_{i1}^{(t-1)}$. Gaps occur when a freshly paired $w_{i1}^{(t)}$ is several scans old. The whole process begins with $t = T$ (the last scan), setting $F^{(T)} = P^{(T)}$, $W^{(T)} = \emptyset$ and, consequently, $W_1^{(T)} = \emptyset$.

The mechanism advancing the scanning bar causes jitters of a few hundred microns that make it difficult to exactly map position (x, y) in one image of a plate to the same position in a subsequent image (Fig. S2-1). Thus, for two positions to correspond to each other, we require them to be within distance r. In optimizing the assignment f, we therefore exclude (penalize) pairings between a position in $W_1^{(t)} \cup F^{(t)}$ and a position in $P^{(t-1)}$ whose Euclidean distance is larger than r. This speeds up the optimization process and tunes it to detect stationarity, since pairings between more distant positions are unlikely to refer to the same animal in our time-lapse context.

Various sources of noise and measurement error may produce non-optimal behavior of the locomotion analysis procedure. For example, $w_i^{(t)}$ may drift perceptively over time, as plates dry slightly or the scanner bar drifts. The r-criterion may allow two time-tracks to drift close to each other and "cross over", or failure of the SVM model to detect worms may produce time-tracks with excessive intermittency or very long gaps. Spurious events may be introduced by fast moving worms that are coincidentally caught close to the anchor point of a worm time-track. To handle

all of these cases, paths are discarded or merged ("cleaned up") according to a panel of criteria, prior to being submitted to local movement analysis.

- A w_i is discarded if it is too short (total duration of less than 6 hours), too sparse (i.e.it contains points at only half of all its observation times), or it drifts quickly (the average distance between consecutive scans is greater than 5 pixels).
- If two or more w_i are located very close to each other (less than 50 pixels) near the beginning of one w_i and near the end of the other, it is likely that a single stationary animal was subjected to a one-time displacement, causing it to be incorrectly split into two w_i. If the corresponding ends of the two w_is do not not overlap for a very long time (more than two hours) and are not separated by a large gap in time (more than twelve hours) the two w_is are merged.

The end result of this analysis is a set of time-tracks, as shown in Fig. 2f of the main text. Each straight track corresponds to a stationary worm with an associated stack of cropped images.

S5.2 Posture analysis of stationary worms

The input to local movement analysis is a series of cropped images $X_i^{(t)}$, $t = 1, \ldots, T$ of a stationary worm i, with $t = 1$ indicating the image from the first scan and $t = T$ the image from the last scan. For the sake of less clutter, we shall omit the index i with the understanding that the following analysis pertains to a single time-track.

All images have undergone background subtraction as described in Note S2 section S2.1. Furthermore, a scanner does not produce perfectly aligned images in consecutive measurements of the same plate (Fig. S2-1). Each image $X^{(t)}$ is therefore aligned computationally with the consecutive image $X^{(t+1)}$ to remove machine jitter. An image is a set of pixel intensities on an (x, y)-grid and represents a discrete sampling of the continuously varying intensities being measured. Although pixels are located at integer offsets, the scanner alignments errors resulting from the advancing scanner bar can vary continuously. To account for such sub-pixel errors in alignment, we sub-sample images to allow registration offsets of distances less than a single pixel.

A typical stationary worm series with posture changes is shown in Fig. 2 of the main text. The

objective of local movement analysis is to detect these changes by comparing consecutive images of the same stationary worm. Since animals may take a variety of postures, and various mutants may present novel postures, we do not attempt to determine the exact form of these changes. Instead, we simply determine the overall change Δ_t in pixel intensity inside the alignment frame between $X^{(t)}$ and $X^{(t+1)}$. This strategy also has the advantage of not relying on accurate segmentation, allowing subtle head movements to be detected even in the common situation in which heads are not fully retained during segmentation.

Consecutive images of the same plate region may vary in overall intensity simply because of variations in the output of the scanner light source or stray light. To compensate for differences in average brightness between subsequent images, we use as the local movement score

$$\Delta_t = \sum_{i=1}^{N} |(x_i^{(t)} - \overline{X^{(t)}}) - (x_i^{(t+1)} - \overline{X^{(t+1)}})| = \sum_{i=1}^{N} |x_i^{(t)} - x_i^{(t+1)}| - |\sum_{i=1}^{N} x_i^{(t)} - \sum_{i=1}^{N} x_i^{(t+1)}| \qquad (7.3)$$

where $\overline{X^{(t)}}$ denotes the mean pixel value in image $X^{(t)}$ and the summation runs over all N pixels in the alignment of images $X^{(t)}$ and $X^{(t+1)}$. The above expression yields a time series of local movement scores Δ_t, $t = 1, \ldots, T-1$ for a stationary worm (see Fig. 2i in the main text and Figs. S7-1a, S7-1b, S7-3c).

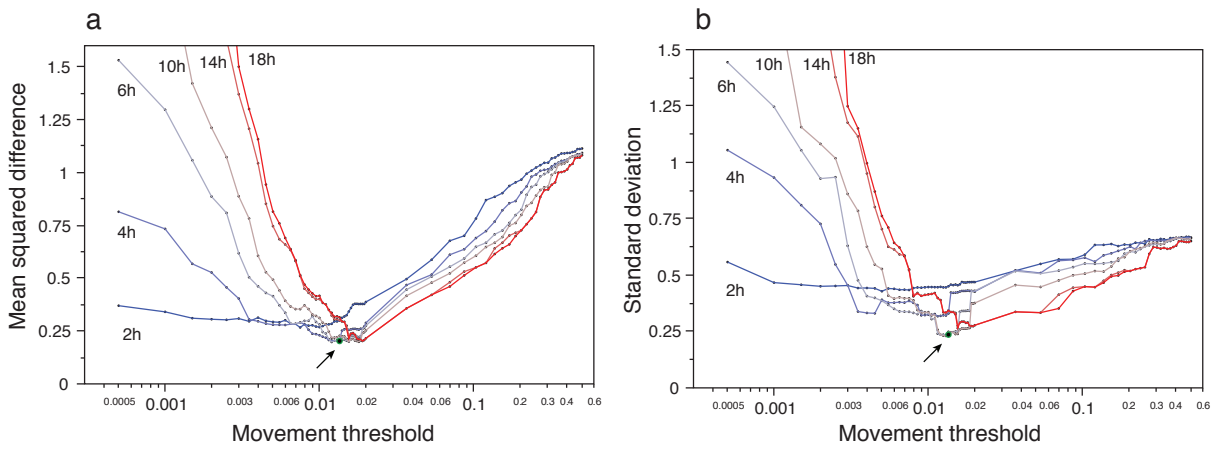

Figure S5-1: Parameter robustness of posture analysis. To determine an optimal combination of the parameters $\Delta_{\max}$ (movement threshold) and τ (hold time), image strips of several hundred animals were inspected using the Worm Browser and their death times were annotated. Many combinations of threshold and hold time values were systematically tested. Each combination was used to estimate a death time for each animal using our posture analysis algorithm, resulting in an estimation error when compared to the by-hand annotation. The panels show the mean squared error **(a)** and the standard deviation of the squared errors **(b)**. The algorithm produces accurate and robust results for threshold values between 0.01 and 0.02, and for hold time values between 6 and 14 hours. The chosen parameter combination is shown as a green dot (arrow).

Additional forms of scanner noise may cause Δ_t to deviate from zero even for dead worms (Fig. 2i in the main text and Fig. S7-2b). This necessitates a one-time calibration step to define a threshold value $\Delta_{\max}$ below which we consider an animal as non-moving. However, dead worms might appear to sporadically move as live animals push them around. We therefore define an animal as *permanently* non-moving, i.e. dead, when it is observed below $\Delta_{\max}$ for longer than a threshold value τ (which exceeds the maximum resting time of live animals). To obtain these two threshold values, $\Delta_{\max}$ and τ, we assemble a "gold standard" time series for several hundred worms whose death times are annotated by visual inspection of the image stack in the Worm Browser (Note S4 section S4.2 and Supplementary Video 6). The thresholds are chosen such as to minimize the deviation between computed and certified death times of the gold standard set, Fig. S5-1. Fig. S5-1 also indicates that the death calls are fairly robust with regard to the so-chosen parameters. Using these parameters, the automated analysis of posture changes returns the first t at which Δ_t last falls below threshold for the required duration of time. The discrete time interval associated with the death event is $[t, t+1]$, and corresponds to the wall-clock time elapsed

between onset of adulthood and scans t and $t+1$.

We examined several image transforms, such as equalization or normalization, prior to pairwise comparison of images that might amplify and clarify local motion. Yet, none yielded a distinctive advantage over the simple brightness adjustment, equation (7.3).

Supplementary Note S6 — Statistics of worm movement

The movement analysis described in sections S5.1 and S5.2 can be used to partition the population of worms into movement classes, similar to those defined in [Herndon et al., 2002]. Stationarity (or posture) span refers to the time between cessation of translatory locomotion and death, when the only motion consists in subtle changes of head and tail position without affecting overall location. For wildtype and *age-1(hx546)* animals shown in Fig. 4 of the main text, we calculated the stationarity span and find that wildtype animals spent 1.27 ± 0.17 days, whereas *age-1(hx546)* mutants spent 3.0 ± 0.15 days in this state, see Fig. S6-2.

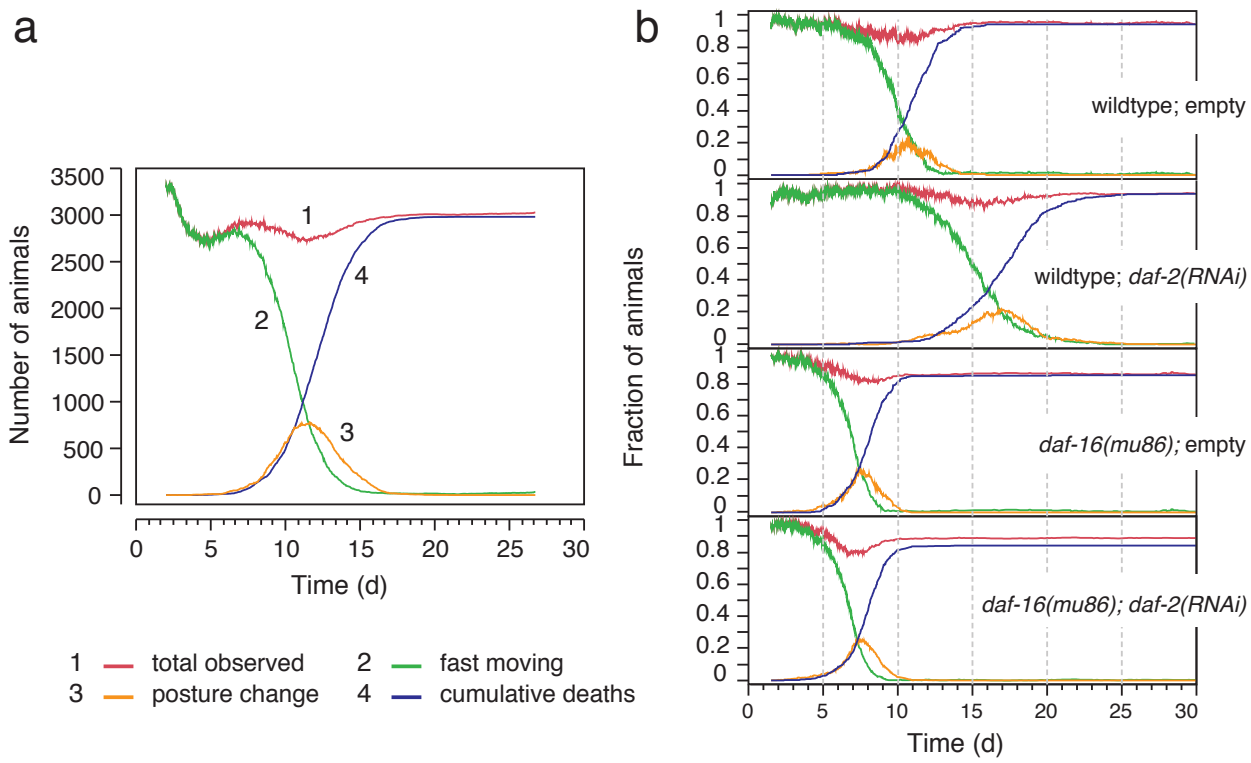

Figure S6-1: Population statistics of movement classes. **(a)**: The red curve (curve #1) reports the total number of wildtype worms observed by the machine at any time in this experiment. (This number changes for reasons detailed in Note S8 section S8.2 and Fig. S8-3). The green curve (#2) is the number of individuals that crawl across the agar, i.e. are not stationary in the sense of Note S5 section S5.1. The orange curve (#3) shows the number of worms that are stationary, but still change posture in the sense of Note S5 section S5.2. The blue curve (#4) reports the number of animals that have died. The overall plot exhibits a population dynamics of movement categories reminiscent of an irreversible chemical reaction chain in which A $\rightarrow$ B $\rightarrow$ C. **(b)**: As in panel (a), but for wildtype and *daf-16* backgrounds with daf-2 RNAi and empty vectors.

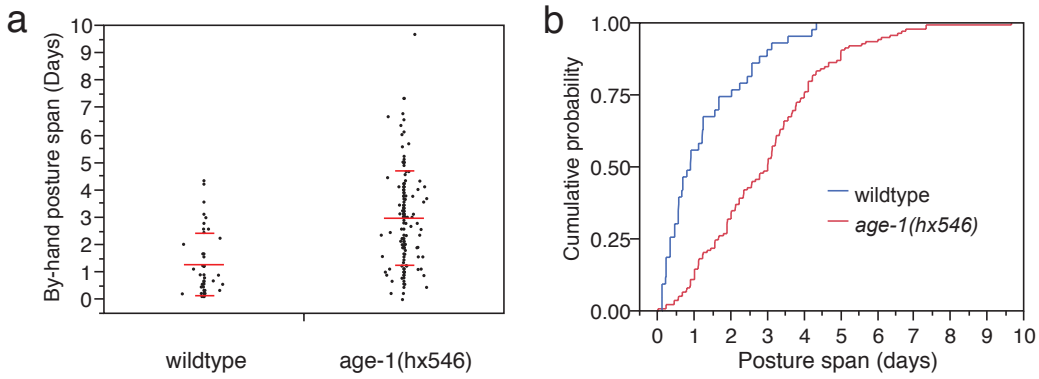

Figure S6-2: Stationarity (posture) span. **(a)** The time spans between cessation of translatory motion and death ("stationarity spans") are shown with overlaid averages (wide bar) and standard deviations (narrow bar). **(b)** Comparison of the cumulative stationarity span distributions. 1.0 minus this curve would be a "survival curve" whose terminal event is death and whose initial time point $t = 0$ corresponds to cessation of translatory motion.

Supplementary Note S7 — A “death phenotype”

After an animal ceases to crawl, the image analysis of the Lifespan Machine produces a time series of registered images of that animal. This image stack, typically used to quantify local movement and identify death times (Note S5 section S5.2), can also be used to longitudinally monitor morphological

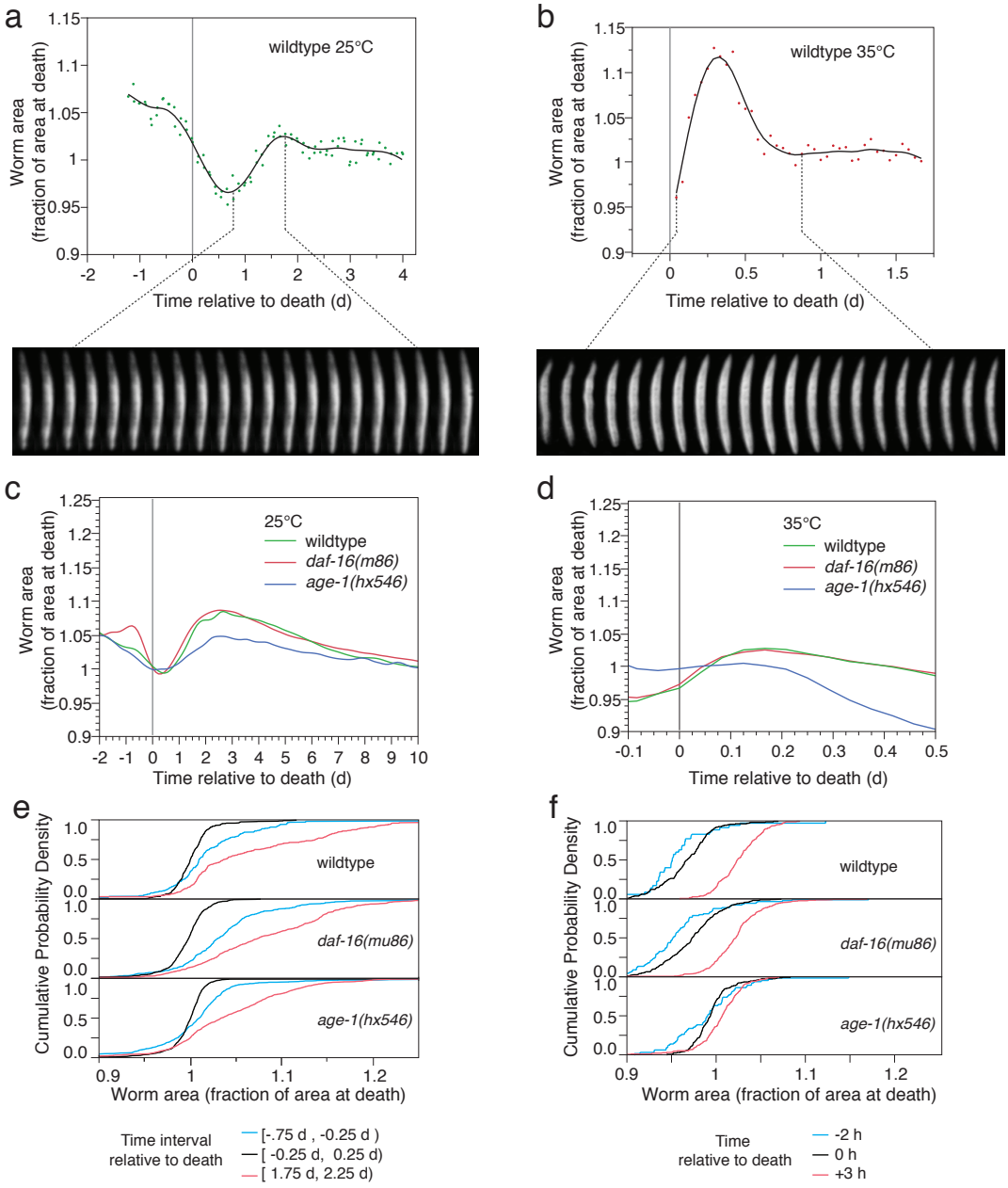

Figure S7-1: The volume of a worm dips at or near death. **(a)** We followed the total area (in pixels) of a single wildtype animal dying on OP50 at 25 °C. This area is plotted against time before and after the animal's death ($t = 0$) as determined by visual inspection. Dots represent scores; solid lines are cubic spline fits [Reinsch, 1967, Sall et al., 2012]. The size of the animal decreases starting about a day before its apparent death, after which the trend sharply reverses and size increases. The volume of a worm dips at or near death. (Figure on previous page.) **(b)** A single wildtype animal dying at 35 °C exhibits a consistent increase in size after its death. Caption continued on next page.

Figure S7-1 (Cont'd): Data prior to death is scant, because the stationary movement phase at this temperature is very short. **(c)** To evaluate whether death-related contraction and expansion of worm area is typical, we determined individual time-series of areas for 44 wildtype, 90 *daf-16(mu86)*, and 143 *age-1(hx546)* animals whose death time on OP50 at 25 °C was determined by visual inspection of the image stack using the Worm Browser. A cubic spline was fit to minimize the mean squared error across all individuals at each measurement. This reveals a conspicuous contraction before death and expansion thereafter. **(d)** We proceeded as in (c) for 189 wildtype, 304 *daf-16(mu86)*, and 194 *age-1(hx546)* animals on OP50 at 35 °C. Animals dying from acute heat stress were not observed to show a decrease in size immediately before death. This either represents a true difference in late-life morphological changes compared to animals dying at 25 °C, or the shrinking close to death occurs while animals are still crawling and can, therefore, not be measured with this technique. **(e)** The average behavior of worms shown in (c) and (d) may mask individual variation. To estimate the effect of such variation, we grouped all measurements made within three half-day intervals—the first between 0.75 and 0.25 days before death, the second within ± 0.25 days of death, and the third between 1.75 and 2.25 days after death. The cumulative distribution of each group is shown for each genotype. In all cases, the middle interval (± 0.25 days surrounding death) showed the smallest median and mean size, indicating that animals reach their smallest area just as they cease to move. **(f)** At 35 °C, animal area was considered at three single time points, two hours before apparent death, at apparent death, and three hours after apparent death. Animals appear to begin increasing in size before their last movement, and the size continues to increase afterwards.

features of worms nearing death. This led us to discover that most animals exhibit a stereotyped morphological change at or near the time of their final posture change. They first shrink by 10% or more and subsequently expand by more than that amount. The phenotype is established in Fig. S7-1, its coincidence with the cessation of all spontaneous motion (the death criterion of our method) is shown in Fig. S7-2, and Fig. S7-3 indicates its generality.

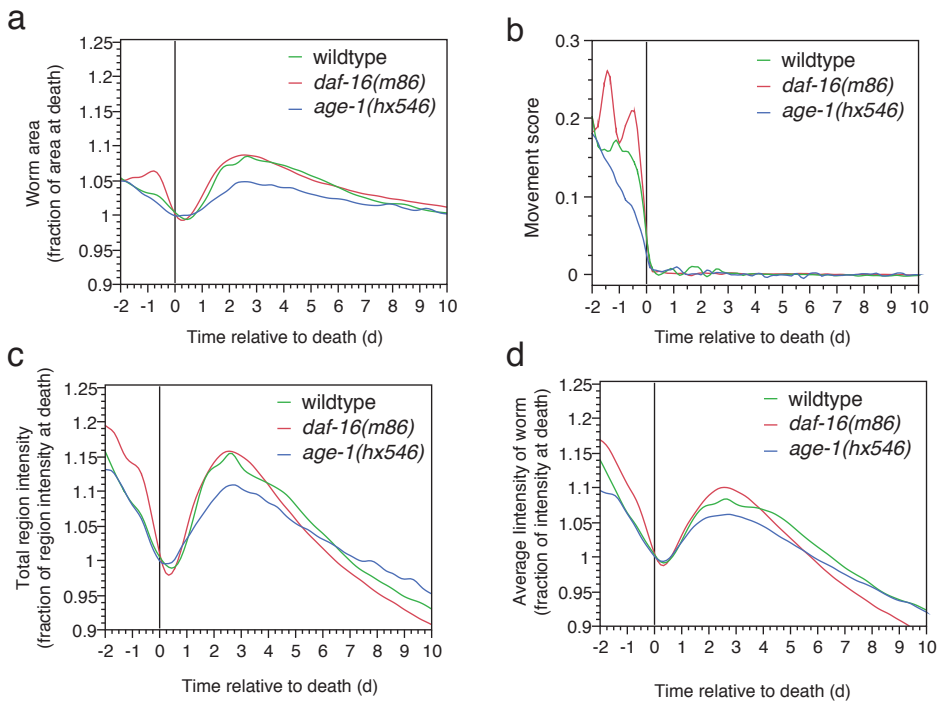

Figure S7-2: Coincidence of volume dip and cessation of motion. Caption on next page.

Figure S7-2 (Cont'd): Coincidence of volume dip and cessation of motion. **(a)** This is the same panel as in Fig. S7-1c. **(b)** To corroborate that these volumetric changes occur contemporaneously with the cessation of movement, we plotted spline fits (as in Fig. S7-1) of the corresponding movement score time-series. Death time identification by movement score is seen to agree with visual determination of death. On average, animals indeed cease to contract and begin to expand approximately at the time of their last movement. **(c)** On visual inspection, animals appeared to darken slightly with size. Pixel intensity of animals and animal area might be related, since area is calculated from a segmented image, whose segmentation depends on pixel intensity. To demonstrate that worms exhibited a net decrease in pixel intensity independently of segmentation, we measured the total pixel intensity within a box of constant size bounding the animal across its image time series. This metric shows the same trend as worm area, providing further evidence that morphological changes surrounding the death time are independent of our segmentation algorithm. **(d)** We determined the average intensity of pixels inside worms over time. This metric, too, shows the same trend as worm area, suggesting that the animals both shrink and diffract less nearing death, while getting larger and darker post mortem.

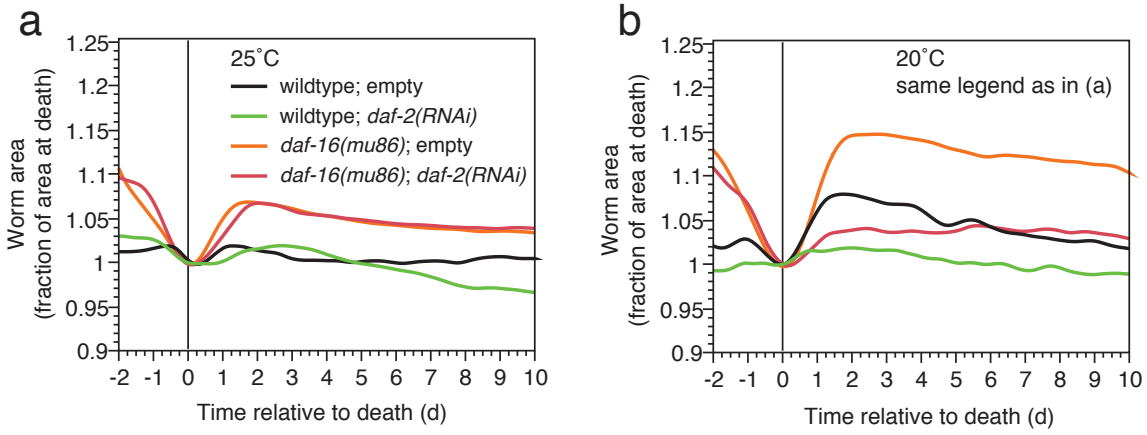

Figure S7-3: Generality of the volume dip. To assess the generality of the late-life characteristics described in Fig. S7-1, we performed the same analysis on animals grown on a different bacterial food source, HT115. **(a)** Area vs time relative to death on HT115 at 25 °C. **(b)** Area vs time relative to death on HT115 at 20 °C.

Supplementary Note S8 — Censoring

S8.1 Right censoring and interval censoring

Individuals are "lost" from a lifespan experiment when supervening events preclude the observation of the age-induced death of these individuals. For example, in both manual and automated lifespan experiments, worms are occasionally found dehydrated on the plate wall. These individuals are lost from further observation, but cannot simply be ignored during statistical analyses, as doing so would involve consideration of a biased sample. Kaplan and Meier proposed a nonparametric method for estimating survival curves that takes into account the contribution of individuals to the population at risk up until the moment they are lost [Kaplan and Meier, 1958b]. We compile our survival curves according to Kaplan-Meier (KM).

Many statistical packages, like JMP, require time-to-event data to assemble a survival curve.

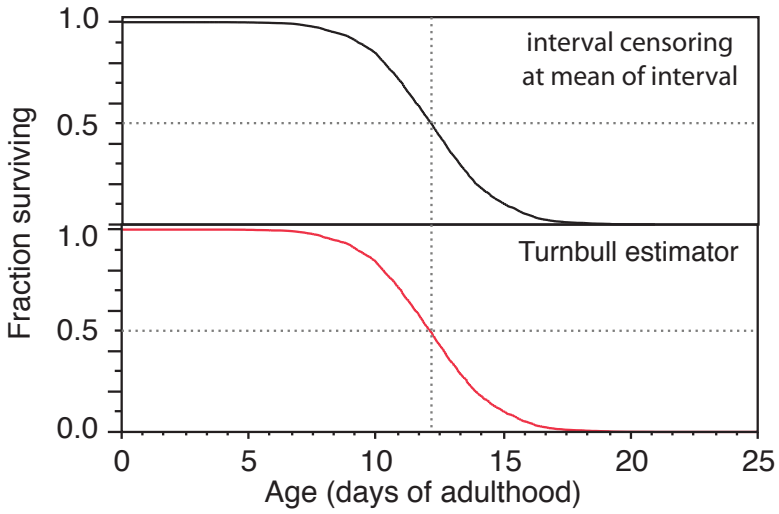

Figure S8-1: Interval midpoint vs Turnbull estimator. See text for details. The two curves are effectively indistinguishable if drawn on top of each other.

Our scans, however, happen at discrete times. Death or loss events are therefore only known to occur within a time interval defined by successive scans. Based on our scanning schedule, these intervals are typically either short (15 minutes) or long (1 hour and 45 minutes), but they can be longer if the system fails to detect a worm in scans immediately preceding its death. Quite generally, when events are only known to have occurred within an interval, a number of approaches exist to handle what is generally referred to as "interval censoring". One interval censoring technique is the Turnbull estimator [Turnbull, 1976], provided by statistical packages such as JMP [Sall et al., 2012] or STATA [Stata Corporation, 2007]. When scan intervals are small compared to even the shortest lifespan in the population, a reasonable and computationally much faster alternative consists in simply choosing the interval midpoint as the time of death for all events known to have occurred in that interval, and then apply standard time-to-event KM.

We compared the Turnbull estimator with the simple midpoint estimation (Fig. S8-1). Using Turnbull estimation, the median of survival fell between 12 days 3.53 hours, and 12 days 3.60 hours. Using the interval midpoint, the median of survival occurred at 12 days 3 hours. This 36 minute difference is overwhelmed by several sources of experimental error that can produce much larger differences, for example temperature. In view of this negligible difference, the substantially faster speed, and the more useful analytics provided by many statistical packages when using time-to-event KM, we chose the midpoint estimate as our standard practice. All our data, however, are optionally supplied in an interval format suitable for Turnbull estimation.

S8.2 Tracking population size and right censoring strategies

As in the manual lifespan procedure, worms whose lifespans are acquired automatically can be lost before their death is observed. Such animals must be right censored when their observation ends. A classic loss occurs when a worm crawls up the vertical sidewalls of a plate. In the Lifespan Machine, this causes a decrease in the total count of worms observed on that plate. Other events that require censoring are of biological nature, such as the "rupture" of a worm that fails to release fertilized eggs. These events are currently not detected automatically, and are counted as deaths, unless explicitly censored by visual inspection with the Worm Browser during the quality control phase that follows data acquisition and image processing.

One source of censoring events is specific to the automated method. Individuals can aggregate into clusters or clumps of two or more animals, where they are usually either misidentified during image processing as a single worm or ignored entirely (Fig. S2-5). Yet, for living animals, the resultant loss is often only temporary, as these individuals become again observable once they leave a clump. Dead animals inside clumps, as well as animals that leave the field of observation will remain missing permanently. We thus refer to an animal that is temporarily unaccounted for as *missing*, to distinguish it from being *lost* (i.e. "permanently missing"). For the purpose of adjudicating (right) censoring events, a worm inside the Lifespan Machine can therefore be in three states—missing, lost, or dead. A significant number of animals may be missing at times. In the manual procedure, observers can search for missing animals and physically separate clumps, but the Lifespan Machine has no such options. This difference requires additional processing to handle censoring. First, we must track the total number of animals on a plate from which we can infer missing animals. Second, we must determine how many among the missing animals were lost (i.e. never returned from the missing state) and at what times these losses occurred.

Because young worms move large distances between consecutive observations, we cannot directly determine when a particular animal goes missing nor, in case of its return, the time when it becomes observable again. Instead, we apply a tunable heuristic, explained below, to estimate the timing of both events. To estimate the statistical effects of this heuristic on our survival curves, we compare two extreme cases that delimit whatever unknowable histories have occurred in an experiment. The curves derived from these extreme censoring protocols are compared to each other in Fig. S8-5.

Let t index scans. Let $d(t)$ denote the number of deaths identified at time t and $D(t)$ the number of deaths that have occurred up until (and including) t, $D(t) = \sum_{i=0}^{t} d(i)$. Further, let $a(t)$ be the number of worms alive at t and N the total number of worms present initially. At any given time a worm is either alive or it has died sometime in the past (including now) or it is missing:

$$a(t) + D(t) + m(t) = N \tag{7.4}$$

with $m(t)$ denoting the number of worms that are unaccounted for (missing) at time t. Automated

analysis of the image stack acquired during the lifespan experiment yields $a(t)$ and $D(t)$, providing us with $m(t) = N - a(t) - D(t)$. Worms mostly go missing in three ways: (i) A scan does not image a worm properly. (For example, it may have crawled near or up the vertical plate wall.) (ii) A worm is misidentified as a non-worm by the classifier. (iii) A worm has disappeared into a clump of worms where it goes undetected. In all of these cases, a worm may return from the missing and be accounted for at a later time. Therefore, a plot of $m(t)$ versus t typically results in a non-monotonic step function, a schematic of which is depicted in Fig. S8-2 (green curve) and an actual example is shown in Fig. S8-3b.

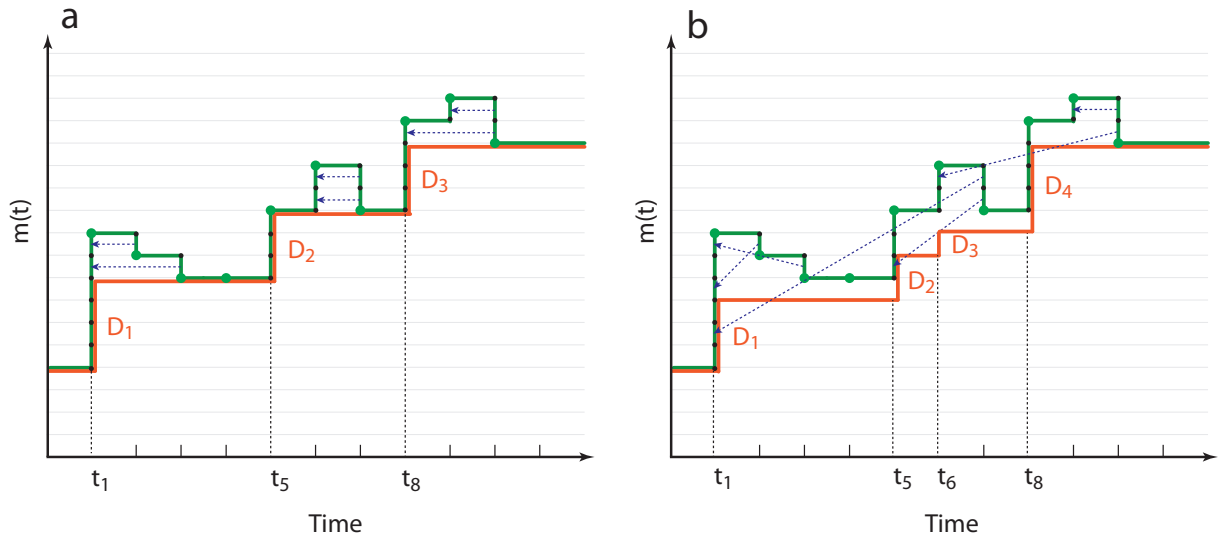

Figure S8-2: Going missing and getting lost (schematic). In both panels, the green curve represents a schematic $m(t)$ of equation (7.4), the number of animals detected missing over time. Each up-step (in the direction of increasing time) indicates the number of animals that have joined the ranks of the unaccounted for. Each down-step indicates the number of those that have become observable again. The orange curve is the censoring curve. Each up-step indicates the number of animals lost from further observation and that must therefore be right censored at that time interval. **(a)**: In this scenario, the orange curve is constructed right-to-left, matching the green curve at first. Each green down-step of size 1 is matched against the closest up-step of size 1 and the orange curve is extended to the green level at that last match. **(b)**: In this scenario, each green down-step is matched at random with a previous green up-step. When all down-steps have been paired, the orange curve is built left-to-right, tracking the number of unpaired green up-steps. Although the green curve $m(t)$ is the same in cases (a) and (b), the two strategies result in quite different censoring curves.

When $m(t)$ decreases, some worms that went missing before t are again accounted for. Yet, we

have no information to deduce when these recaptured animals went missing. In other words, we cannot determine unambiguously how many of the worms that went missing at a given time were actually lost. Consequently, we cannot know exactly how many worms must be right censored *at what time*. The many possibilities are bounded by two extreme scenarios.

In one extreme scenario, we hypothesize that the recaptured animals are those that went missing most recently. This scenario matches each downslope of $m(t)$ with the most recent upslope, see Fig. S8-2a. Each unmatched upslope step then provides the number of animals that were lost to the experiment at that time and must be right censored in the final assembly of the survival curve. This strategy can be formalized as a retrospective (right to left) propagation of the smallest value of $m(t)$:

$$z(t_i) = \min(m(t_i), z(t_{i+1})), \quad i = n-1, \ldots, 1 \text{ with } z(t_n) = m(t_n) \tag{7.5}$$

The series of $z(t_i)$ is a step function that increases monotonically with increasing i. A step at time t_i triggers a censoring annotation in the lifespan log, asserting that $\Delta(t_i) = z(t_{i+1}) - z(t_i)$ individuals are permanently not accounted for and should thus be right censored at time t_i.

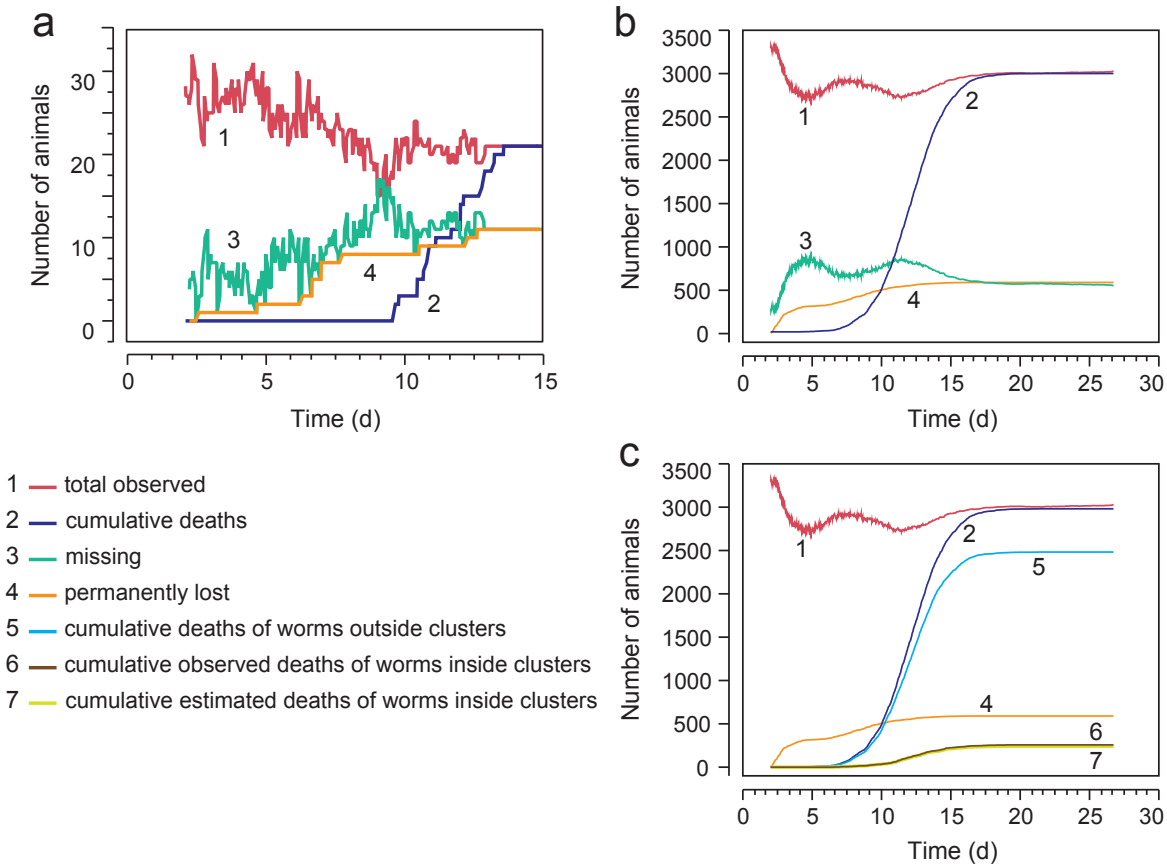

Figure S8-3: Going missing and getting lost (data). In all panels, the red curve #1 shows the total number of observed animals $obs(t)$ at time t, i.e. $obs(t) = a(t) + D(t)$ in equation (7.4). The green curve #3 is $m(t)$, and is obtained by subtracting the red curve from the total population size N at the beginning of the experiment. Determination of N is discussed in section S8.5. The red curve shows change at long time scales, presumably reflecting the overall clustering behavior of worms in the population. Early in the experiment, $obs(t)$ declines because a fraction of worms cluster. Worms then leave clusters, perhaps because food becomes locally scarce. As aging animals become progressively paralyzed, a subset co-locate into juxtaposed pairs or tuples. Close to death, these tuples are misidentified by our image processing as single worms, see section S8.3. The green curve $m(t)$ mirrors $obs(t)$ by construction. The orange curve #4 corresponds to the orange curve in the schematic of Fig. S8-2. **(a)** An example of missing worm detection is shown for a single plate. This plate was chosen because an unusually large fraction of animals was lost. **(b)** For each of 160 plates, the equation (7.5) is applied to individual plates, and the results are aggregated. **(c)** The same data in (b) is replotted with deaths separated into singletons, two-worm clusters, and three-worm clusters.

In the other scenario, starting from $t = 0$, each down-step (recapture event) at time t is randomly matched with an available up-step event ("gone missing") that occurred before t, whereupon that up-step is no longer available for future recaptures. As in the first scenario, unpaired up-steps

correspond to lost animals (see Fig. S8-2b).

In general, we can tune the time horizon within which a recapture event is matched up with a "gone missing" event. Specifically, given a time threshold τ, we attempt to pair a down-step that occurred at time t with an up-step that occurred within $[t-\tau, t]$. If all up steps in that interval have been paired, we choose an up step within $[t-2\tau, t-\tau]$ and so on. The first scenario corresponds to setting $\tau = 1$ (one scan interval) and the second scenario sets $\tau = T$ (the whole experiment).

In all of this, we need to know the total number of worms N at the beginning of the experiment. This is the subject of section S8.5.

S8.3 Worm clusters and interval censoring

The most frequent reason for animals to remain unaccounted for in the Lifespan Machine is the irreversible clumping (clustering) of some worms near the end of their life. In the manual setting, clumping presents no problem, since the experimenter can physically separate individuals that have aggregated. There is no such intervention in the Lifespan Machine.

Early in life, when individuals are highly mobile, aggregation is dynamic and transient, yielding small clusters with odd shapes (such as a Y or T) that our image analysis software can occasionally recognize as multi-worm objects and disentangle computationally (Fig. S2-5). Later in life, when worms become stationary, clusters tend to be more persistent and worms appear to adhere to one another, acquiring an overall linear shape that our image analysis almost always misidentifies as a single worm.

Future developments in image analysis might be able to detect and disambiguate worm clusters. However, at present, the fastest and most reliable alternative is to identify and annotate clusters by visual inspection using the Worm Browser (Note S4 section S4.2). This task can be performed easily and quickly. Processing an experiment with three thousand animals takes less than an hour and is part of the data validation recommended for any novel data set to confirm that all steps of the image analysis pipeline are functioning as expected.

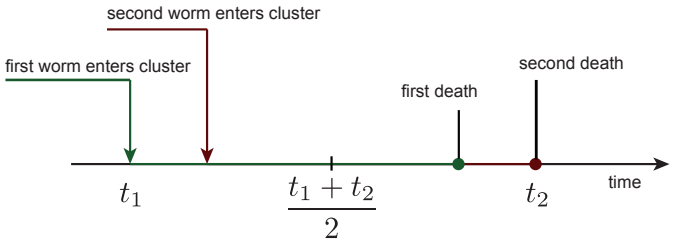

Figure S8-4: Interval censoring of a terminal worm pair (schematic). The first worm becomes stationary at t_1. A cluster is nucleated once a second worm joins. A terminal pair is created if the latter becomes stationary in the cluster. The last of the two animals dies at t_2 (a valid individual death time). The first worm to die unnoticed within the terminal pair is then assigned (upon visual inspection of machine death calls) a death time corresponding to the midpoint of the time interval between t_1 and t_2. To the Lifespan Machine the whole episode appears as a single worm becoming stationary at t_1 and dying at t_2.

Although image processing typically misclassifies a persistent cluster of n worms as a single individual, the death event of a cluster is a valid event associated with the last worm dying. Once a cluster is identified, our software finds the first stationary worm that nucleated the cluster, thus defining a time interval within which the remaining $n - 1$ worms must have died, Fig. S8-4. This, then, is handled by interval censoring, either using the Turnbull estimator or the interval midpoint, see section S8.1.

S8.4 Censoring strategies for worm clusters

Our current image-processing software requires manual intervention to distinguish single animals from pairs or triplets lying in close contact to each at death. To evaluate the statistical effects of such annotations, we compare 6 different strategies for handling multi-worm clusters, Fig. S8-5. All death times were acquired as discussed in Fig. 3 of the main text.

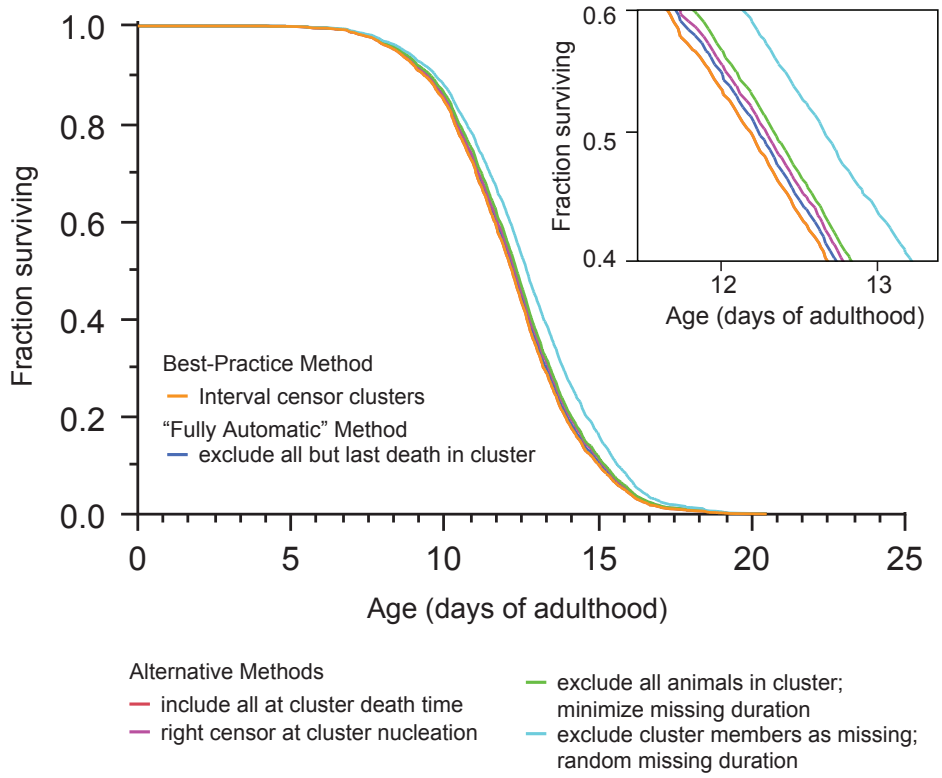

Figure S8-5: Censoring strategies for terminal worm clusters. Curve 1 (orange): Our best practice method, in which multiple-worm clusters are manually annotated using the Worm Browser. The image-processing pipeline determines an interval bounded by the nucleation time of a cluster and its death time. (The cluster is typically misidentified as a single worm.) One individual is assigned the cluster death time, all others are assigned the interval midpoint. This is the curve shown in Fig. 3 of the main text. Curve 2 (blue): No manual annotations are considered, and only a single death time is de facto recorded for all animals in a cluster. This death occurs at the death time of the last animal to cease movement within the cluster. Curve 3(red): Using manual annotations of multi-worm clusters, all members of the cluster are assigned the same death time, which is the cluster death time. Curve 4 (purple): Using manual annotations of multi-worm clusters, worms entering into clusters are declared lost at the time of cluster nucleation and right censored. Curve 5 (green): Using manual annotations of multi-worm clusters, such clusters are completely eliminated from the death record. As a consequence, all animals that disappeared into terminal clusters are handled according to the "missing worm procedure" (which also handles transient clusters), as described in section S8.2. Curve 5 is obtained when the procedure operates according to equation (7.5), which minimizes the time a worm is assumed to have been missing. Curve 6 (cyan): As for curve 5, but the procedure pairs missing events and recapture events at random, see section S8.2.

S8.5 Determining total population size

Sometimes the total number of worms N at $t = 0$ may be unknown; for example when animals are loaded onto plates without being counted. In this situation, we estimate the total population size by determining and comparing two quantities. First, we take the maximum of $a(t) + D(t)$ for $t = 1, \ldots, T$ (see section S8.2 and Fig. S8-3). This is the largest number of worm objects ever identified during the course of the experiment. Let this number be N_0. The visual identification of final clusters with the Worm Browser provides another estimation of the total number of worms, call it N_c. N_c is the number of worms found in clusters at the end of the experiment plus the number of singleton death events $D_s(t)$, as determined by the machine. We take N to be the maximum of N_0 and N_c.

Supplementary Note S9 — Statistical power

The effect of a perturbation on lifespan is established through comparison of the perturbed population against an unperturbed control, using a statistical test. In the case where the perturbation does have an effect, we would like the test to reject the null hypothesis that the two populations are identical not only with a sufficient level of confidence $1 - a$ (where a is the rate of false positives, type I error) but also with a sufficiently high probability $1 - b$ (where b is the rate of false negatives or type II errors). a is called the statistical significance (typically chosen as 0.05) and $1 - b$ is known as the power, typically chosen to be 0.8 or 0.9. We used our actual lifespan data to reason about the sample size required to reach a desired power.

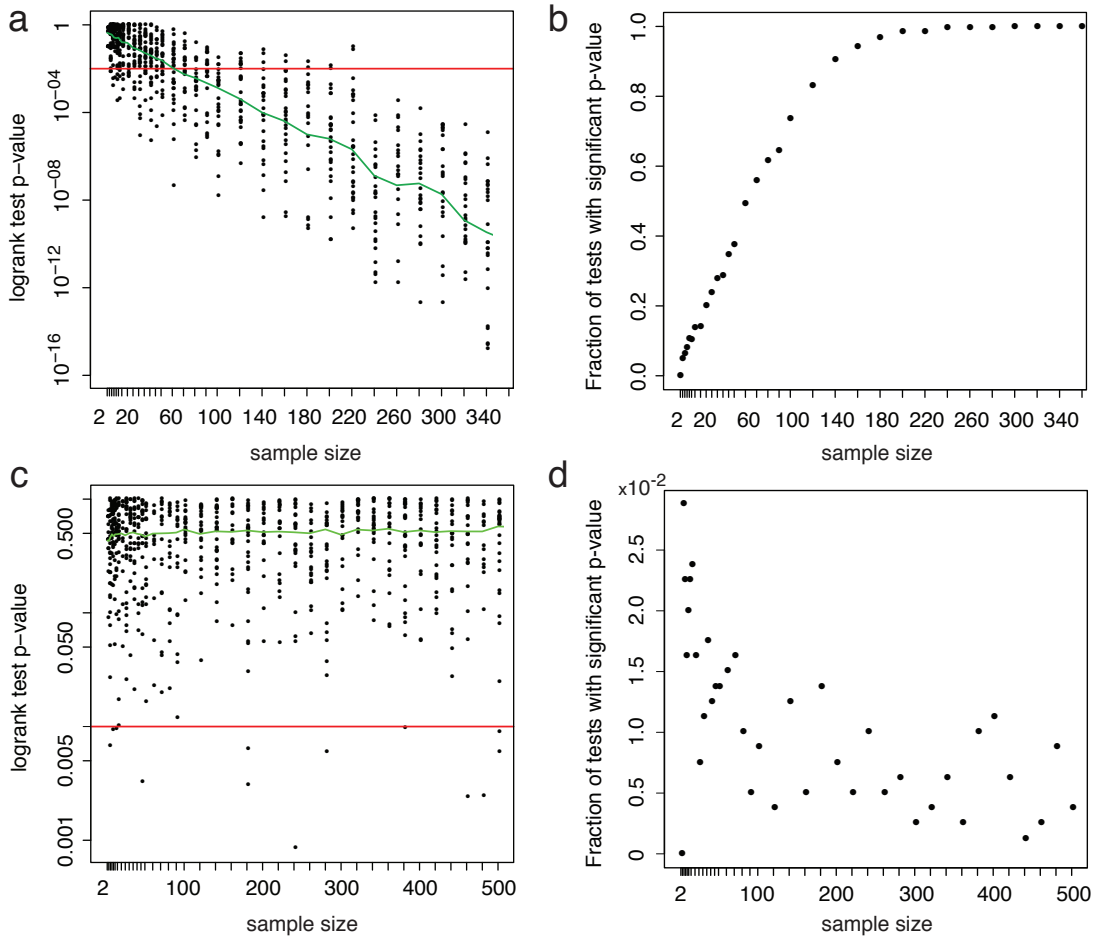

Figure S9-1: Statistical power. The top panels, **(a)** and **(b)**, illustrate the sample size dependence of the Log-rank test. In 800 replicates, we drew two sub-population of the specified size (abscissa) at random from a wildtype and *daf-16* mutant population observed at high temperature (34.5 °C) by the Lifespan Machine. At this temperature the *daf-16* mutation appears to produce a 10% reduction in average lifespan (Fig. 5a). For each pair of random subsets, we used the Log-rank test to decide whether or not to reject the null hypothesis that the samples are statistically indistinguishable. Panel **(a)** shows the scatter of p-values found in each replicate at each sample size. (To allow fast rendering of the figure, only a random subset of all 800 replicates at each size are shown) Note that the ordinate of panel **(a)** is logarithmic. The red line indicates 99% confidence (i.e. 0.01 significance). Caption continued on next page.

Figure S9-1 (Cont'd): Each dot above the red line is a test with non-significant outcome, incorrectly failing to reject the null hypothesis. The green line marks the median of the p-value distributions. Panel **(b)** shows the frequency with which the (wrong) null hypothesis is correctly rejected as a function of sample size. This is known as the power of the test. (The complement of this frequency is the type II error rate or the rate of false negatives.) It can be readily seen how the power increases with sample size. The bottom panels, **(c)** and **(d)**, are controls for a case in which the null hypothesis is true—in this case the samples being compared are drawn from the same (wildtype) population (at 25 °C). Note that in both cases, the abscissa is linear in scale, but tick marks were placed at each size sampled. As for panel **(a)**, the ordinate of panel **(c)** is logarithmic. See text for details.

In one case we compare two samples, one drawn from the wildtype population and one from the *daf-16* population presented in Fig. 5b of the main text (both at 35 °C). In that case the null hypothesis is false by construction. In another case we compare two samples both drawn from the same wildtype population of 3578 individuals (25 °C) presented in Fig. 3b of the main text, illustrating the situation in which the null hypothesis is true. The red line indicates 0.01 significance (99% confidence). Each dot above the red line represents a test correctly not rejecting the true null hypothesis, whereas each dot below the red line corresponds to a test that turned out significant, thus erroneously rejecting the null hypothesis (type I error or false positive). In Fig. S9-1d we plotted the fraction of tests, as a function of sample size, that reject the true null hypothesis. As expected, this fraction is very low even at small sample sizes and approaches zero for large samples. In contrast, Fig. S9-1a reports the case in which the null hypothesis is wrong, as high-resolution data shows *daf-16* mutant populations to differ in mean lifespan from wildtype by about 10% at 34.5 °C. In the simulation, we observed sample size to have a substantial effect on power, the probability of rejecting the wrong null hypothesis at a given level of confidence, Fig. S9-1b. With a population of 110 animals every 5th test will get it wrong (a power of 80% or 20% false negatives). At around 200 animals, power has increased to 99%.

In summary, our arguments in this section indicate that sample sizes of sufficient power to detect with high confidence a difference in survival prospects of 15% or more are comfortably within reach of the manual method. Detecting 10% differences at high confidence and power becomes more tedious (on the order of 200 animals) and quickly becomes a bottleneck for determining the mortality statistics associated with a larger number of environmental or genetic conditions. The detection of small effect sizes also places more stringent requirements on the reproducibility of

environmental conditions, hardly achievable with the routine manual approach. Finally, biological insights often depend on showing the *absence* of an effect, in which case very small effect sizes need to be determined with high confidence, requiring population sizes of several hundred animals. These sizes are routinely achievable with the Lifespan Machine at high throughput.

Supplementary Note S10 — Aggregation of survival curves

S10.1 Device-specific environmental effects on lifespan

The natural units of aggregation in the Lifespan Machine are its "units of confinement": plate, scanner, incubator. The procedures described in the previous sections apply to a single plate and yield a survival curve for that plate. Each plate produces data with high temporal but low statistical resolution, as a plate contains about 30-50 worms. Better statistics can be produced by aggregating multiple plates containing animals of the same strain, either directly with our software or using a statistical package like JMP [Sall et al., 2012]. Building curves for individual plates provides a highly valuable sanity check, as it affords an opportunity for statistical tests of equivalence and for recognizing when entire plates may have to be censored (for example, due to fungal invasion, desiccation, or fogging). Data from individual plates are aggregated to generate survival curves at the level of individual scanners. Before pooling scanner-aggregated curves into an overall curve at the incubator level, they may have to be registered to account for small device-specific environmental differences, especially local temperature.

C. elegans lifespan is sensitive to environmental conditions and lifespan assays can detect the effects on survival that result from small differences in these conditions. It is therefore important to maintain a constant environment across the different plates in an experiment. This is technically difficult using the manual assay, and the problem is compounded when using a large, distributed apparatus like the Lifespan Machine. Factors such as temperature (Fig. S1-3b) or the composition of the bacterial lawn may vary slightly between plates and scanners, creating distinct micro-environments that can impose "extrinsic" lifespan variation on the "intrinsic" variation of interest. Scanners, for example, can differ in temperature because they produce and dissipate their own heat within a large confined space (incubator) whose airflow they obstruct. We see temperature differences of up to 1°C (Figs. S1-3b and S1-3c) between scanners located inside the same incubator, and identify statistically significant differences between the lifespans of animals housed on different scanners (Fig. 3c of the main text and Fig. S1-2). Since the populations on

each scanner are drawn from the same homogeneous population, the differences between survival curves most likely reflect temperature differences rather than biological variation.

We deploy fans (Fig. S1-1) and optimize scanning frequencies (Figs. S1-3b) to minimize temperature differences between scanners. Ideally, these modifications would reduce environmental variation to the point where its effects fall below the detection limits of our method. This, however, is a high standard for our equipment, since the Lifespan Machine can monitor large populations at short intervals, enabling it to detect very small effects from environmental variation. We therefore pursue a different strategy: detecting and correcting such environmental variation computationally by using a linear regression model in which scanners are treated as values of a categorical variable ("device"). This model makes a few assumptions: that each scanner has the same effect on the lifespan of all individuals it measures and that it alters lifespan linearly, either shifting it by a constant offset or scaling it by a constant factor. The latter case is known as the Accelerated Failure Time (AFT) model.

S10.2 The device-effect regression model

Let t_{ij} be the lifespan of the jth individual on scanner i, $i = 1, \ldots, s$. The effect of environment on lifespan can be modelled in two simple ways. First, additive effects might result in time shifts between survival curves, such that t_{ij} is $\epsilon_{ij} + \beta_i$ with β_i a constant offset specific to scanner i and ϵ_{ij} (a "residual") the corresponding lifespan absent the "treatment" by scanner i. Second, multiplicative effects might result in a proportional scaling of lifespans, such that t_{ij} becomes $e_{ij}\beta_i$ (the AFT model). For the sake of brevity, we describe additive and proportional effects using the same (additive) model, with the understanding that in the proportional case the t_{ij} denote the *logarithm* of lifespan.

The model can be written as

$$t_{ij} = \alpha + \beta_i + \epsilon_{ij} \quad \text{if } i \neq s \tag{7.6}$$

$$t_{sj} = \alpha - \sum_{i=1}^{s-1} \beta_i + \epsilon_{sj} \tag{7.7}$$

In these equations, α is a landmark against which the device-dependent effects are determined. A

possible choice is the grand mean μ of all N lifespans (or their log) measured in the same incubator or the average lifespan of a population on a particular scanner. The grand mean is a good choice when the environment varies evenly between scanners, and worms are distributed evenly across the scanners. When α is the average lifespan on a specific scanner with temperature T, we effectively correct the lifespans on all other scanners as if they were at temperature T. The $\beta_i, i = 1, \ldots, s-1$ are the deviations of the mean lifespan $\mu_i = \sum_j t_{ij}/N_i$ of the N_i individuals on scanner i from α and ϵ_{ij} is the deviation of t_{ij} from μ_i. Since we must have $\sum_{i=1}^{s} \beta_i = 0$, β_s is determined by $\beta_s = -\sum_{i=1}^{s-1} \beta_i$.

The relation with linear regression is made explicit by introducing a categorical variable $\vec{X}_j$ whose (vector) value encodes the scanner hosting individual j:

$$t_j = \alpha + \vec{X}_j \vec{\beta} + \epsilon_j \tag{7.8}$$

with $\vec{\beta}$ the vector of scanner effects to be determined. Estimates of the parameters of these types of equations are handled routinely with statistical packages such as R [R Development Core Team, 2009] or JMP [Sall et al., 2012].

In our case, we also must account for the fact that we place animals on scanners several days after onset of adulthood. The time that individual j is exposed to the environmental impact of a scanner is therefore not its lifespan t_j, but $t_j - t$ or $\log(t_j - t)$, where t is the age at which animals were placed on scanners.

The Cox model constitutes an additional possibility to account for a scanner effects. In Cox regression, the scanner environment is assumed to have an additive effect on the log of the hazard function. We refrain from running a Cox model, which requires fitting a baseline hazard, for the purpose of the present paper.

Any of the AFT-class of regression models can be applied to censored data using the Buckley-James estimator based on modified least-squares estimating equations [Buckley and James, 1979]. This method replaces each censored observation by the conditional expectation given the observed data and the covariates. Because the conditional expectation itself depends on β, the estimating equations must be solved iteratively.

Supplementary Note S11 — Hazard

S11.1 Hazard analysis

The hazard rate is a fundamental statistical observable in aging. It is the instantaneous rate at which individuals that survived to time (age) t die in the next instant. The role played by this rate in a non-reproducing population, like a plate of sterile worms, is analogous to that of a rate constant in a first order chemical reaction that thins out an initial population of molecules by degradation. However, the hazard rate $h(t)$ is typically not a constant, but a time-dependent quantity that may (and often does) increase with time reflecting the functional decline associated with aging: $ds(t)/dt = -h(t)s(t)$, where $s(t)$ is the survival function. $h(t)$ can be observed directly by counting the frequency of death events relative to the survivors (the population at risk) at the beginning t of intervals of fixed duration Δt ("period hazard"): $h(t) = [s(t) - s(t + \Delta t)]/\Delta t s(t)$. To do this accurately, however, requires population sizes that are larger and more frequently observed than is typically the case in manual lifespan experiments. The Lifespan Machine makes this fairly straightforward, and the machine can be scaled up, if desired, to test many conditions.

We determined three kinds of parametric hazard functions for our data: Gompertz, Weibull and Log-logistic. Gompertz and Weibull hazards represent exponential and polynomial increases in the risk of failure, respectively, and are widely used to describe mortality data in biological and technological systems. The Log-logistic is a Weibull distribution with a deceleration term. We used it here as a reference to compare the deceleration of mortality rates observed in most cases. Such deceleration has been described anecdotally before in *C. elegans* [Vaupel et al., 1998, Johnson et al., 2001, Baeriswyl et al., 2009] and flies [Curtsinger et al., 1992, Carey et al., 1995] (and humans, wasps and automobiles [Vaupel et al., 1998]), although never in stress regimes (e.g. high temperature) nor systematically for a large number of mutants.

The parameters of model hazard functions—Gompertz, Weibull, and Log-logistic—were obtained by maximum likelihood estimation (MLE) using the lifespan data (death times) of the wildtype

population shown in Fig. 3 of the main text. Gompertz and Weibull hazards do not decelerate and MLE estimates of the parameters of these distributions will yield a poor fit. The origins of the deceleration phenomenon deserve an empirical study in their own right, which we leave to future work (but see [Weitz and Fraser, 2001b, Steinsaltz and Evans, 2004, Steinsaltz and Evans, 2007] for computational and theoretical approaches). In the present data analysis we focus on the functional form of early mortality, which is a natural and easier first task in organizing our volume of data. To this end—and unless otherwise noted—when obtaining MLE estimates of model parameters, we only use lifespan data up to the median, right-censoring the remainder and thus eliminating the late phase in which hazard decelerates. This is what we mean in legends or captions when referring to, for example, "50% Gompertz". The phrase "100% Weibull" then means that the full range of lifespan data was used, except for cropping 1% at the noisy tails. The maximum likelihood estimation of model parameters was performed using the `flexsurv` package of `R` [R Development Core Team, 2009].

A quantile-quantile (Q-Q) plot is a diagnostic for assessing the extent to which survival data conform with the distributions whose parameters we determined. In a Q-Q plot we graph the time it takes a functional form (Gompertz, Weibull, or Log-logistic) to reach a certain quantile against the time it takes the empirical survival function to reach the same value. In our figures we place the former on the abscissa and the latter on the ordinate. The values (plotting positions) are determined by the steps in the Kaplan-Meier survival curve obtained from our data (Note S8 section S8.1). The 95% confidence intervals in the Q-Q plot derive from the confidence intervals generated in the survival curve assembly from data (Greendwood's formula) using the `survfit` function of `R`'s `survival` package. The extent to which empirical data and (MLE) fitted functions coincide can be assessed by how closely the points hug the identity line. It is worth pointing out that while we determine parameters based on the data up to the median lifespan, we extend the Q-Q plot to include the entire range of data. This makes the onset of deceleration neatly visible as an upward departure from the diagonal for Gompertz and Weibull hazard functions.

S11.2 Hazard of wildtype

The next two figures evidence that high-resolution lifespan data for wildtype conform with a Weibull distribution better than a Gompertz distribution up to times slightly beyond median lifespan when deceleration becomes noticeable.

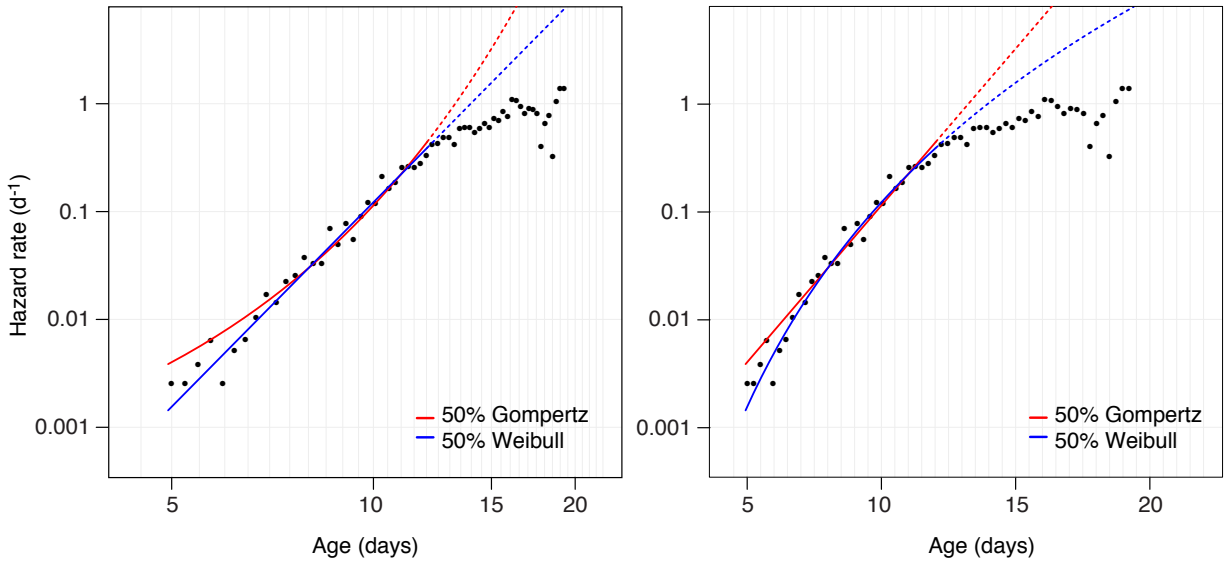

Figure S11-1: Wildtype high-resolution hazard functions at 25 °C. In this figure we graph, for the purpose of visual comparison, our hazard rate data (black dots) and the Gompertz (red) and Weibull (blue) model hazard functions as determined by MLE from our lifespan data. The black dots are produced by subdividing the duration of the experiment into equal intervals and reporting the average risk of death (the number of deaths as a proportion of the population initially at risk) within each interval. The comparison with model hazard functions suggests that a power-law (Weibull) describes early mortality better than an exponential (Gompertz). The left panel reproduces Fig. 2d of the main text and is plotted using a log-log scale (in which the Weibull hazard is a straight line), while the right panel uses a log-lin scale (in which the Gompertz hazard is a straight line). The hazard fits are drawn solid over the data range used in the fits (i.e. up to median lifespan) and dotted elsewhere.

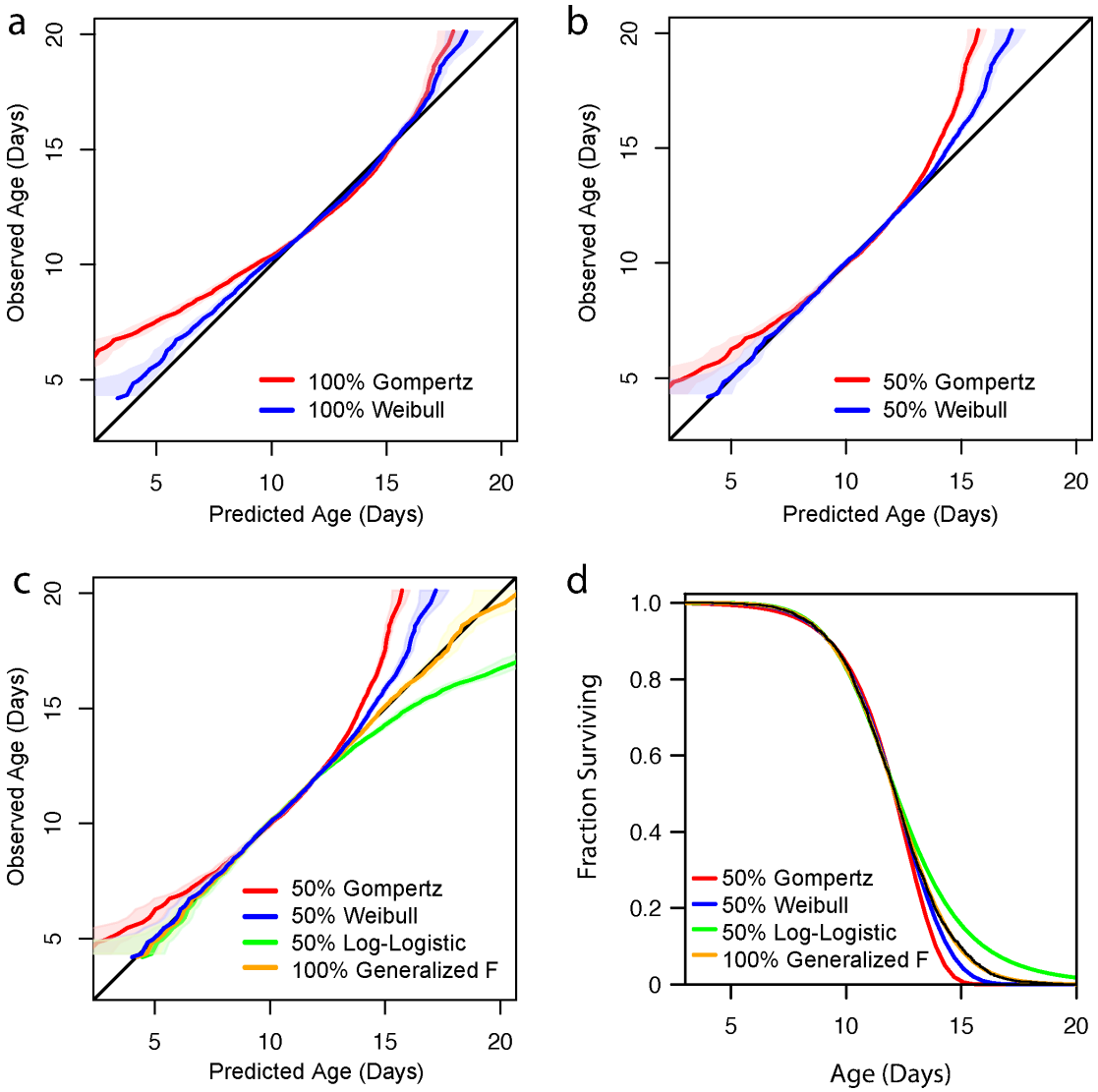

Figure S11-2: Wildtype high-resolution quantile-quantile plots at 25 °C. The quantile-quantile plots (see section S11.1) visualize the extent of (dis)agreement between data and a variety of models. **(a):** The parameters of Gompertz and Weibull hazards were determined by MLE using the full range of data (except for the 1% crops at the tails). Clearly, the agreement is poor, since the hazard deceleration present in the data is at odds with functions that cannot decelerate. **(b):** Model parameters were determined using data only up to the median lifespan. This panel is a striking indication that early mortality in wildtype *C. elegans* follows a Weibull more closely than a Gompertz. **(c):** Two further models are considered: a Log-logistic, whose parameters were determined using data up to the median, and a generalized F distribution [Cox, 2008] that takes into account the whole range of data. The Log-logistic coincides with the Weibull distribution at early times, but evidently overestimates the extent of hazard deceleration. The generalized F is a four-parameter umbrella that encompasses both Log-logistic and Weibull distributions. Not surprisingly, it can be tailored to the data very well. Panel **(d)** shows the survival curves implied by the fitted model parameters. The black curve with the narrow 95% confidence interval represents the non-parametric Kaplan-Meier curve based on our data. The generalized F sits snug underneath. The main differences to the other models are chiefly in the tail. Again, the blue Weibull curve is seen to match the data well beyond the median.

The hazard analysis of the manually scored results also indicates a better fit to the Weibull than the Gompertz distributuion.

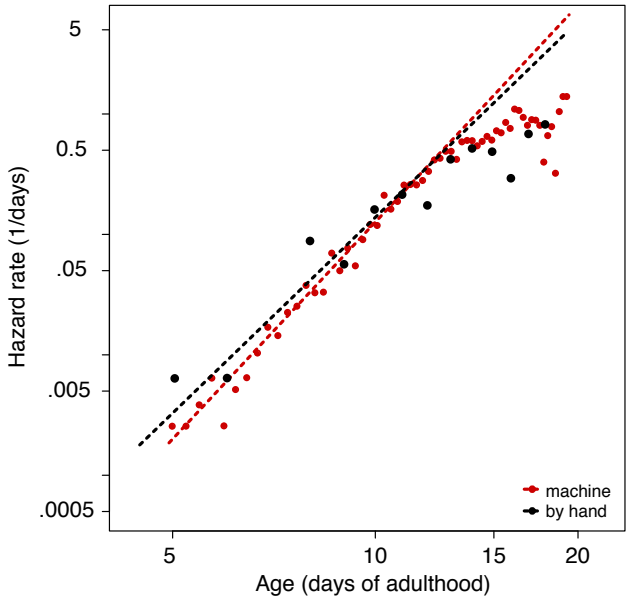

Figure S11-3: Wildtype hazard at 25 °C scored manually. This figure compares the empirical hazard of the manually scored wildtype population (Fig. 3 in the main text) with that of the LM-scored population. The Weibull hazard is seen as the better fit for early mortality data obtained with either method.

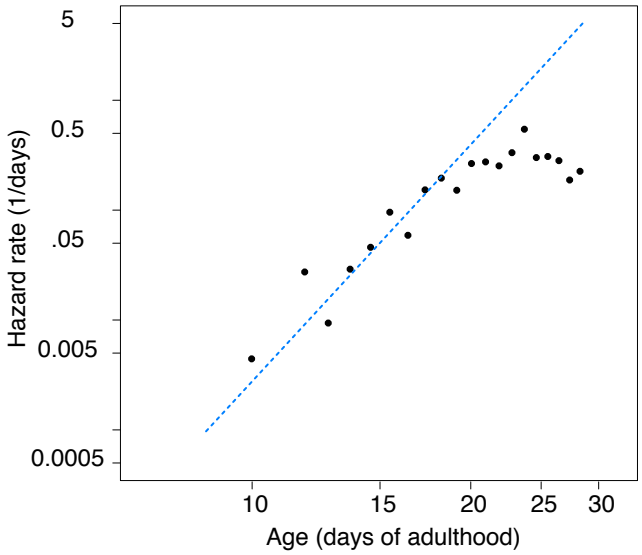

Figure S11-4: Wildtype hazard at 20 °C on HT115. This figure depicts a hazard rate plot with Weibull fit obtained from a wildtype population (220 animals) feeding on HT115 bacteria at 20 °C. It shows that the general features of the mortality kinetics at 25 °C persist at 20 °C.

S11.3 Hazard and sample size

In Fig. S11-5 we demonstrate the impact of population size on hazard estimates. The effects of population size demonstrates the precision afforded by large populations.

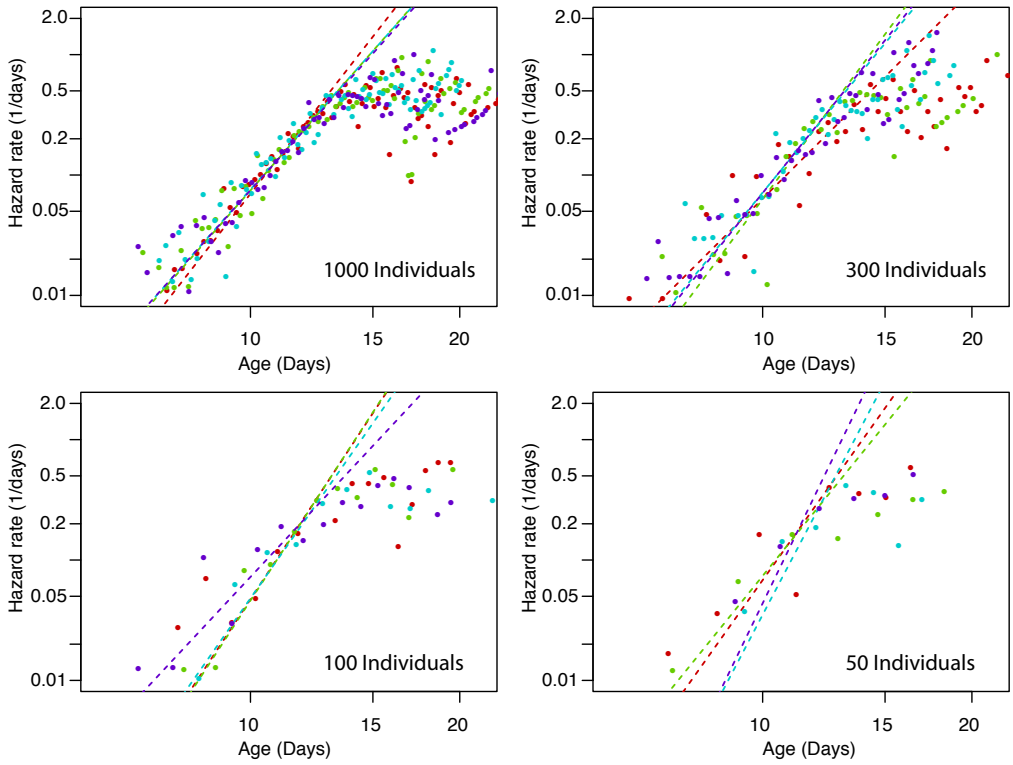

Figure S11-5: Population size and hazard parameter estimates. The four hazard panels on the left show empirical hazard data (dots) and Weibull fits (MLE parameter estimates to original lifespan data) for four random samples (colors) at the indicated sample size. This conveys a sense of how sample size impacts the estimation of hazard parameters.

Supplementary Note S12 — Control for scanner effect

C. elegans researchers define death in terms of macroscopically observable proxies. The most widely used proxy is the failure to move in response to prodding with a platinum wire. Our automated approach uses a slightly different proxy—the permanent cessation of all spontaneous motion as determined by retrospective analysis of long-term observations. We observe these two proxies correlate surprisingly well (Fig. 3b and Fig. S12-1). Yet, absent a principled theory of death, it seems conceptually impossible to "control" for one proxy by comparing it to another. Specifically, we cannot expect the comparison of a manually performed lifespan experiment to act as a strict control for an automated experiment, as the two measure somewhat different phenotypes. This is why we established quality and consistency of our method in two ways: First, by evaluating the Lifespan Machine's ability to recapitulate in a self-consistent fashion known effects on lifespan resulting from a wide variety of interventions, such as mutations, RNAi, and exposure to stressors of physical (heat) and chemical (t-BuOOH) nature (Figs. 3, 4, and 5 of the main text). Second, by developing software facilitating the human *ex-post* inspection and validation of the visual record associated with an experiment.

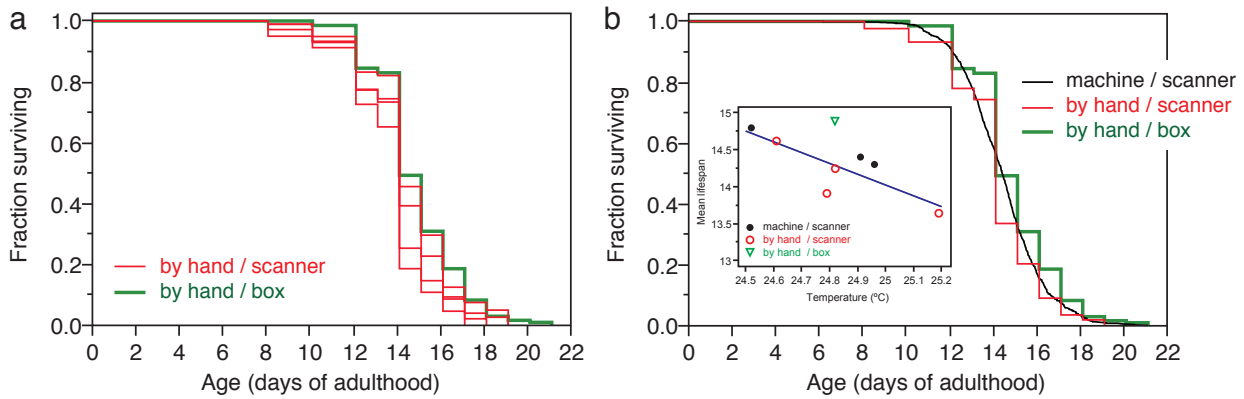

Figure S12-1: Absence of a "scanner effect". To test whether the scanner environment affects the lifespan statistics of worms, we manually scored plates residing on scanners, thus exposed to the same environment as plates processed by the Lifespan Machine. Panel **(a)** compares a set of such manually acquired survival curves with curves assembled from the traditional manual scoring of plates stacked in a box inside the same incubator. Both curves agree. Panel **(b)** compares the two manual approaches (plates inside a scanner or in a box) with the Lifespan Machine. The inset of panel (b) shows the temperature differences between animals on individual scanners, monitored by various techniques, inside the incubator set at 25 °C (see also Fig. S1-2). All plates were seeded with HT115(DE3), since the data were acquired in conjunction with Fig. S13-3.

We sought to determine what effects the scanner environment (light, vibrations, etc.) might have on nematode lifespan. For example, a stimulation of worm movement in response to light has been noted using standard microscopy techniques [Ward et al., 2008]. Mathew et al. [Mathew et al., 2012] suggest that the light produced by flatbed scanners may elicit a similar response. We transferred several thousand wildtype animals onto plates stored in two environments—inside operating flatbed scanners and inside a box sitting inside a neighboring, temperature-matched incubator. For animals stored inside flatbed scanners, we observed a portion by hand and the remainder using the Lifespan Machine, Fig. S12-1.

We first discuss animals scored by hand. We observed that animals located inside scanners lived 18 hours shorter than those housed outside scanners in a separate incubator. To place this difference in the context of temperature, we identified a significant correlation between measured scanner surface temperature and mean lifespan, both through linear regression on plate means ($R^2 = 0.65$; $p = 0.027$) and in Cox regression with temperature as a continuous covariate ($p = 0.002$). Using our linear regression as a predictor for lifespan we found that the difference

between scanner and box groups was reduced to 15.6 hours, or 4.4% of lifespan, representing our best estimate of the effect of the scanner micro-environment on nematode lifespan. We suspect that effects of this magnitude, despite our attempts at accounting for the effect of temperature, might nevertheless result entirely from limitations in our ability to accurately measure temperature as it varies between scanner and box environments.

We then considered animals housed on scanners, scored either by hand or the Lifespan Machine. We found that animals observed by hand appeared to live 6.2 hours shorter than those observed by machine, or 1.8% of total lifespan. This close correspondence between our method and the manual procedure suggests that our proxy for death, cessation of movement closely matches that of the manual, and that any worm that is capable of spontaneous movement does eventually move.

Supplementary Note S13 — Additional survival curves

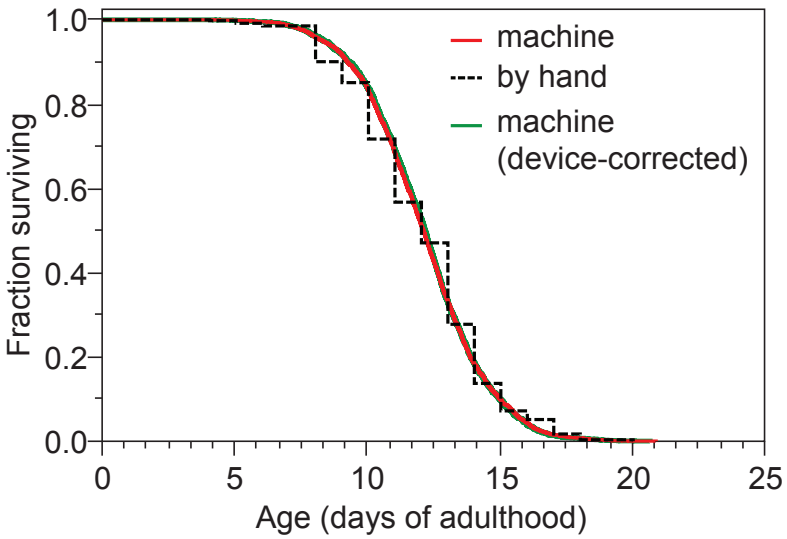

Figure S13-1: Wildtype at 25°, device-corrected. In Fig. 3 of the main text, panel 3b shows the wildtype survival data aggregated from the entire experiment and panel 3c shows survival data from individual scanners with their device corrected versions using the categorical regression of Note S10 section S10.2. This figure shows, for completeness sake, the aggregated device-corrected curve alongside the curves of panel 3b for comparison. At the aggregate level, device correction is seen to have only a minor impact in this case, suggesting that the small variations in temperature across scanners and their effects nearly cancel out.

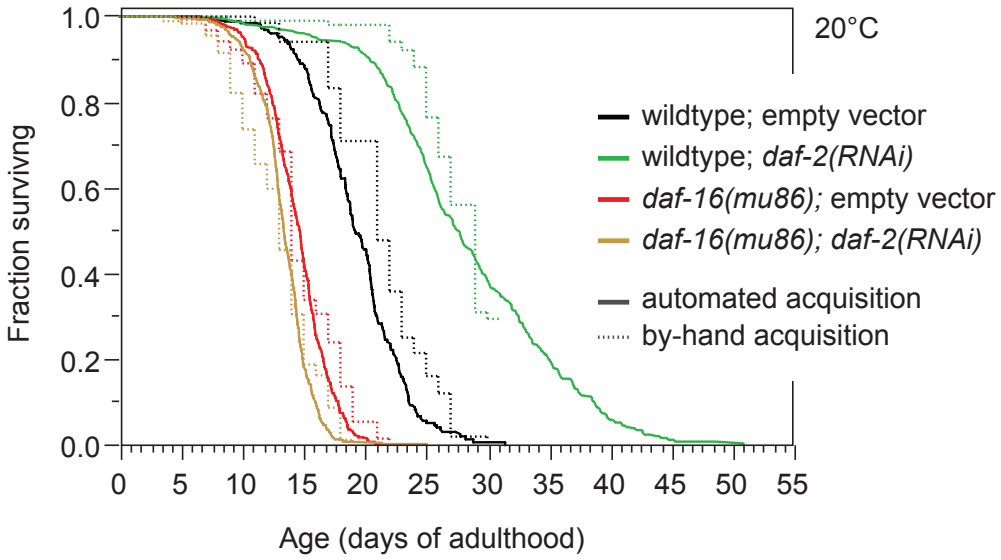

Figure S13-2: RNAi Lifespans at 20 °C. To test the performance of the Lifespan Machine at 20 °C, we placed a subset of animals prepared for Fig. 4a in the main text on scanners in an incubator calibrated at 20 °C. The by-hand control for wildtype on *daf-2(RNAi)* was truncated as experimenters left for winter holidays; the Lifespan Machine continued unaffected.

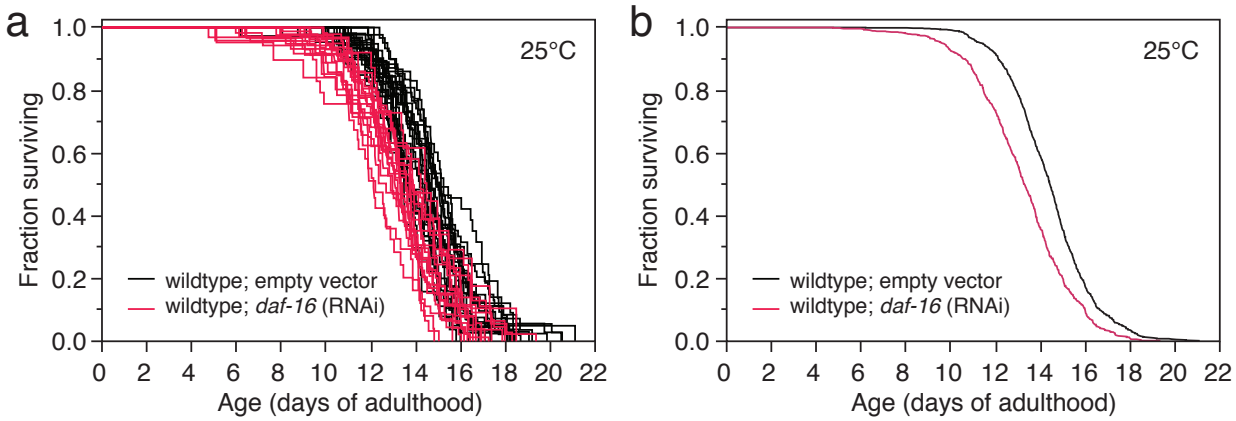

Figure S13-3: *daf-16(RNAi)* at 25 °C. RNAi of daf-16 has been reported to shorten lifespan [Larsen et al., 1995, Morley and Morimoto, 2004]. Here we note that the RNAi knockdown is noticeably weaker than the mutant shown in Fig. 4c of the main text. Panel **(a):** survival curves by plate. Panel **(b):** aggregated survival curves.

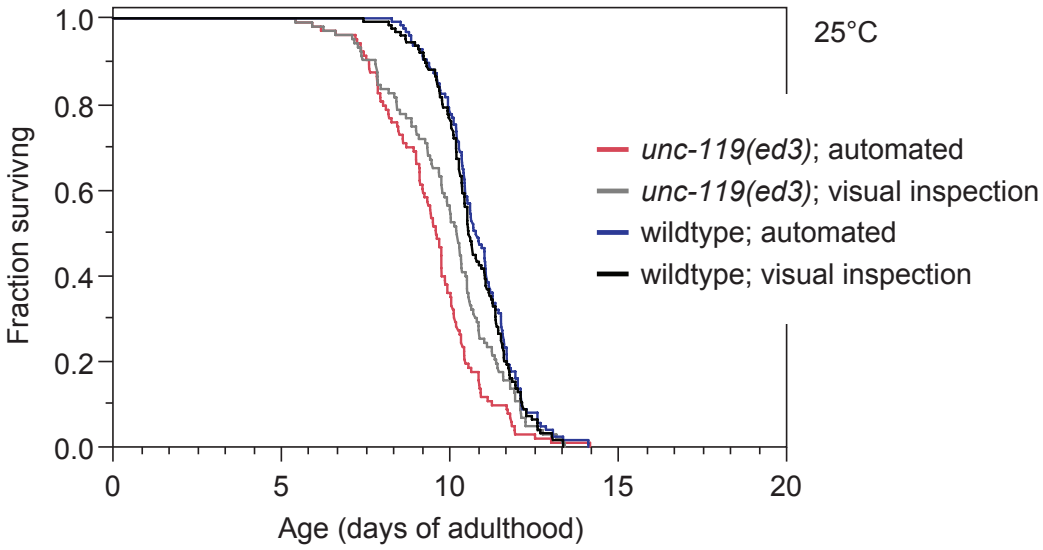

Figure S13-4: Automated lifespan assay of a severely uncoordinated mutant. We used *unc-119(ed3)* animals to assess the performance of the Lifespan Machine on worms with unusual movement phenotype. These animals show abnormal development of the nervous system and exhibit both slow and uncoordinated movement [Maduro et al., 2000]. Using the posture analysis parameters optimized for wildtype animals, our image analysis consistently identifies death times of *unc-119* as being, on average, only 1.5 days shorter than wildtype. The death times of *unc-119* were slightly underestimated by 0.3 days in mean lifespan compared to the visual inspection. This difference is not statistically significant ($p = .09$). Movement parameters can be optimized for specific behavioral mutants, if necessary.

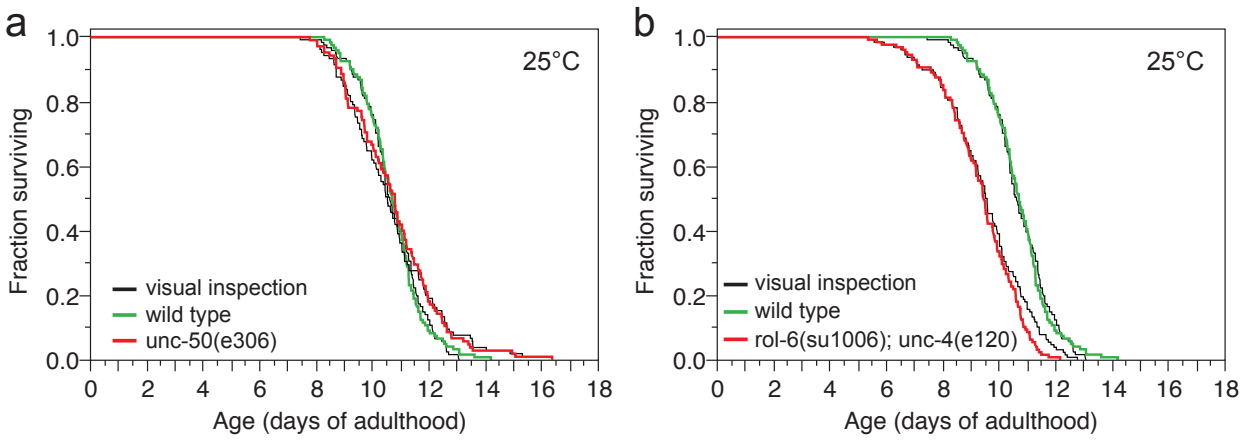

Figure S13-5: Survival curves of *unc-50(e306)* and *unc-4(e120);rol-6(su1006)* at 25 °C. Panel **(a)**: Hermaphrodites possessing the unc-4(e120) allele have been previously characterized [Gems and Riddle, 2000] as having a significantly shorter lifespan than animals with the wildtype allele. Pairwise comparisons between automated analysis (colored curves) and visual inspection of the image record (black curves) are provided as in Fig. X of the main text. Panel **(b)**: The inclusion of a strain exhibiting a roller phenotype further demonstrates the broad applicability of our approach.

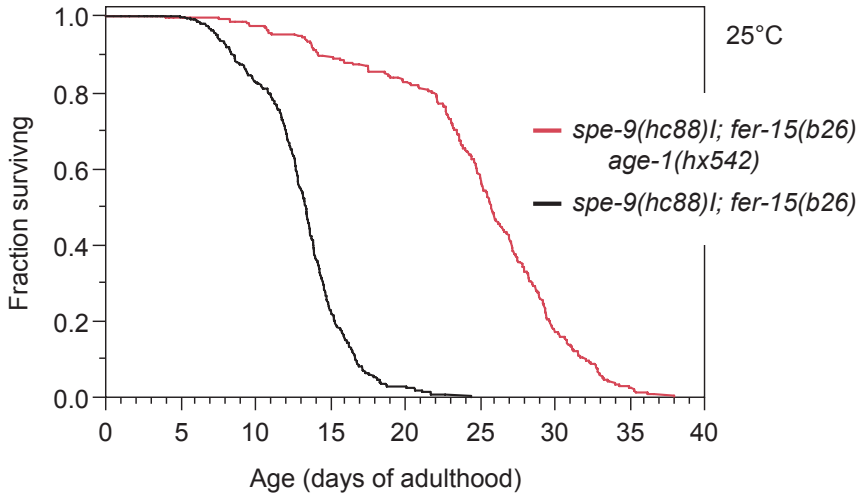

Figure S13-6: Lifespans of temperature-sensitive self-sterile mutants. To avoid the potential for cohort confusion, lifespan experiments with self-fertilizing hermaphrodites on the same plate must prevent the accumulation of live progeny. This is often achieved with 5-fluoro-2′-deoxyuridine (FUdR) [Hosono, 1978]. To assess the performance of the Lifespan Machine without the use of FUdR, we assayed two strains containing temperature-dependent sterile mutations in the *spe-9* and *fer-15* genes. Placing the eggs of such mutants at 25 °C results in sterile populations.

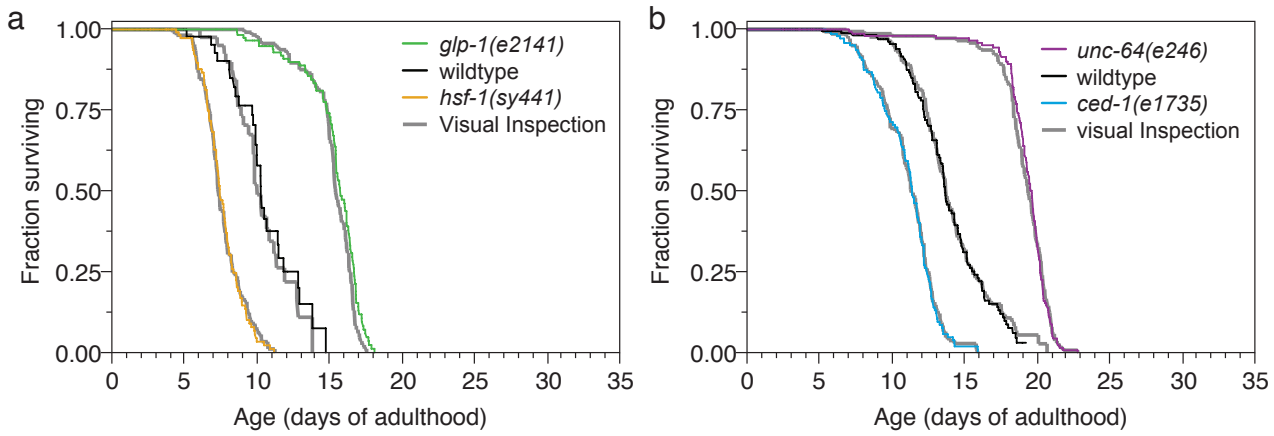

Figure S13-7: Visual controls of additional mutant survival curves. Fig. 4 in the main text misses the juxtaposition of the survival curves acquired by machine with those obtained by visual inspection with the Worm Browser for the mutants shown here in panels **a** and **b**. Visual and automated curves were indistinguishable by Log-rank ($p > 0.04$). *glp-1* and *hsf-1* strains were grown at 15°C before being shifted to 25°C.

Supplementary Note S14 — Resolution and scalability

The apparent smoothness of a survival curve depends not only on population size but also on sampling (scanning) frequency. A large population that is observed infrequently will yield a survival curve with large steps, as many deaths accrue between successive measurements. An example is the wildtype manual control presented in Fig. 3b of the main text, where a population of ≈ 500 animals is observed at a frequency of once a day, producing large steps near the mean lifespan. At the other extreme, frequent observation of small populations also yields survival curves with pronounced steps, as in that case each individual death accounts for a sizeable fraction of the population. Evidently, some advantages of large populations are only fully realized when paired with frequent measurements. Therein lies one advantage of an automated method like the Lifespan Machine. Another advantage is scalability. When the overall lifespan of a population is short—as it is, for example, at $35°$C (Fig. 5 of the main text)—even our baseline observation frequency of 24 measurements per day permits relatively few observations, potentially eliminating the advantage of large population sizes. In such a case, however, we can increase the effective temporal resolution by using multiple scanners and staggering their scan times to effectively monitor at all times some sub-population of the total population residing in the apparatus. We use this technique to improve the temporal resolution of 35 °C survival assays, resulting in smoother aggregated survival curves (Fig. 5b of main).

Supplementary Note S15 — Use cases

The lifespan machine captures images of worms using flatbed scanners, whose data are analyzed and integrated by software. Scanners can be easily added and removed, allowing the technique to be applied at small scale in installations including one or two scanners (as in Fig. 4e–h of the main text, where four mutants were monitored in two rounds on two scanners), or at large scale (as in Fig. 3b, where one mutant was monitored across ten scanners) As such, the method reduces labor in a variety of experimental contexts; adding scientific value by enabling new or old (and recalcitrant) questions to be approached with commensurate data.

- *One or two scanners*

  A single scanner holds 16 plates with 35 worms each for a total of 560 individuals. As indicated in Note S9, this capacity is sufficient to detect 10% lifespan differences between two genotypes at 99% confidence and 99%power. Therefore, a single scanner enables standardized, high-resolution survival assays, appropriate for detecting subtle genetic effects. A single scanner installation involves three components—an incubator (for temperature control), a scanner (for image capture), and a computer (to control the scanner and store data). The incubator is standard equipment in most labs working with *C. elegans*, and small installations may simply use one shelf of a pre-exisiting incubator.

  The use of single-scanner installations constrains the number of different populations that can be simultaneously monitored and the statistical power at which those populations can be characterized, limiting the number of possible applications. For example, the characterization of mortality-rate deceleration may not be feasible using a single scanner, since hazard deceleration becomes prominent only after the median lifespan (at which point the dwindling population size attainable with single scanners may provide insufficient statistical power). Moreover, we find several instances where the proportional hazards assumption of certain statistical tests (e.g. Fig. 4b) are not appropriate, necessitating the use of less powerful statistical approaches requiring larger populations to be monitored. All of these limitations are compounded by the loss of worms (Note S8 section S8.1) from plates, as well as the loss of entire plates due to fogging or desiccation, which often lowers the

population sizes captured on a single scanner from their theoretical maximum by about 10

- *Five or Ten Scanners*

  The limitations of single-scanner installations are easily alleviated by using multiple scanners. The lifespan machine is designed to allow multiple scanners to be deployed with little additional difficulty or complexity, compared to setting up a single scanner. If a laboratory acquires a dedicated computer and incubator for automated lifespan analysis, it is not very laborious to fill the incubator with scanners. A large incubator with 5 shelves will hold 10 scanners, forming a natural "medium" sized installation. Scaling up from a single scanner to ten scanners is merely a matter of purchasing, modifying, and plugging in the additional scanners; our image server software is designed to accept new scanners even while experiments are running.

  A ten-scanner installation affords population sizes that provide greater statistical power not only for standard analyses such as the Log-rank test or Cox regression, but also for the application of somewhat more exotic analyses, such as non-proportional-hazards approaches or techniques that directly estimate the time-dependent hazard function. Our intended purpose of providing high statistical power is not to simply permit detection of increasingly small lifespan effects, as there will be a point when statistical significance ceases to correlate with biological significance. Instead, the machine is meant to enable investigations into the shape and functional form of the hazard rate, which often requires the evaluation of subtle effects that occur late in life, at which time effective population numbers have dwindled, as well as early in life where only a fraction of deaths in each population occur. Installations of five to ten scanners allow the experimenter to perform multiple replicates of these types of analyses.

  Installations of five to ten scanners also permit experimenters to screen through libraries of mutants and RNAi constructs. When population sizes of 100 animals are chosen, three strains can be reasonably tested per scanner, allowing thirty strains to be measured in a single incubator. Consider that a wildtype population should be run on every scanner along with the target populations. This will necessarily curtail real estate for mutants that can be co-located on the same scanner. One should also consider that this use captures entire

survival curves for each population, increasing the quality but necessarily decreasing the throughput compared to the single time-point assays currently used in many high-throughput screening contexts.

Stress resistance assays, such as the thermotolerance experiments shown in Fig. 5 of the main text, also benefit from an incubator full of scanners. Being able to spread animals across multiple scanners allows observation times to be staggered, thereby permitting a finer temporal resolution for capturing the time-dependent hazard function.

- *Twenty or more scanners*

  Construction of a scanner farm requires the dedicated use of several incubators or a temperature controlled room. At this scale, any number of projects becomes possible, including further development and testing of the apparatus itself in parallel with running experiments. The tunable statistical resolution makes our method ideally suited for large rapid reverse-genetic and chemical screens for effects on survival at intermediate or higher statistical resolution.

Supplementary Note S16 — Limitations

As a system, the Lifespan Machine is best evaluated in terms of the quality of the survival statistics it produces. Yet, components such as imaging hardware, worm classifiers, and movement detection algorithms, can also be evaluated in isolation. For example, adoption of imaging devices other than flatbed scanners (e.g. cell phone cameras) may decrease both the physical bulk of the Lifespan Machine and the potential for temperature variation between plates. Furthermore, we find that our SVM classifiers produce the best results on younger worms, strains with larger bodies, and plates with thinner agar lawns, as these situations yield more distinct images of worms. Our posture analysis algorithm works best on animals that do not exhibit long pauses in late-life movement, as our use of a pause threshold to remove noise works less well in these cases. Occasional imperfect performance in all these components can be seen by scrutinizing Supplementary Videos 1–5. We expect that improvements will be made in many components of the Lifespan Machine. For example, movement time-series analysis might use Hidden-Markov models, or worm detection strategies might integrate image data from consecutive frames. The consequences of such local improvements, however, should be evaluated with regard to both the independent behavior of the component and its integrated effect on the overall quality of the survival curve.

Beyond transient limitations that are more reflective of the current state of development of the Lifespan Machine, our approach is restricted in some more fundamental aspects.

- *Worm density on a single plate*

  Our software can analyze 100 or more worms on each plate. In practice, however, worm behavior introduces complexities that limit the usefulness of our analysis of high-density populations. At all densities, we see animals often die closely as pairs, juxtaposed head-to-head and tail-to-tail. This circumstance makes it hard for the image analysis to *automatically* distinguish a single dead animal from a pair of dead animals stuck together (Note S8 section S8.3) These cases we resolve through rapid visual inspection using the Worm Browser (Note S4 section S4.2). For many users, it may be permissible to instruct the survival curve assembly software to discard such pairs entirely, as doing so does not appear

to cause problems of misestimation, see Fig. S8-5, presumably because such "paired dying" occurs randomly. At high densities, however, we see an increasing number of animals dying in groups of four or more, at which point even careful human inspection cannot disentangle the varoius death times. Ultimately, we prefer to minimize the potential scope of such problems by maintaining low animal densities. We find that with 35 animals per plate the frequency of worms dying in close juxtaposition is very tolerable. Lowering worm density can further reduce issues associated with worm aggregation, if desired.

- *Deaths from causes unrelated to aging*

  The current version of the Lifespan Machine cannot discern causes of death that are unrelated to aging, such as bagging. The risk of bagging (and perhaps of non-aging deaths in general) decreases as egg production slows with age. As a result, bagging occurs almost exclusively during a nematode's reproductive window. In our by-hand data sets the frequency of such events is less than 5% and in some cases 0%, agreeing with previous characterization [Apfeld et al., 2004]. Bagging and vulval ruptures can be identified in the visual data validation step using the Worm Browser, where they can be annotated for censoring. In Fig. S8-5, we demonstrate how various strategies for handling a much more frequent censoring event—the disappearance of worms—have measurable but ultimately negligible effects on survival curves. We therefore do not expect typical bagging rates to have much of an impact on the quality of survival data produced by the Lifespan Machine. However, certain environments and specific mutations may increase the rate of bagging, rendering automated analysis infeasible. In such cases we recommend the following strategy for recognizing (and removing, if desired) bagged animals. Because bagging occurs primarily early in adulthood when deaths are very infrequent, animals need not be placed onto scanners until late in their reproductive span. Populations can be manually inspected for bagging (or any other visible phenotype) for several days of adulthood, after which they are transferred into the automated imaging apparatus for survival analysis.

- *Inaccessibility of plates and animals during an experiment*

  A clear limitation of our approach is that animals cannot be handled during automated observation. It follows that nematodes must be sterile during observation, so that progeny

do not obscure the analysis of parents. One common means for ensuring sterility is to add FUdR to media. If FUdR is problematic (affecting, for example, the robustness of specific RNAi effects), there are several options available to the experimenter, such as the use self-sterile mutants (see Fig. S13-6). If the experimenter can tolerate some loss of precision during the first few days of death, animals may be passaged by hand as young adults and then loaded onto scanners after they cease laying eggs during mid-life (as in the bagging case described above). If the research community as a whole decides to move away from FUdR, new methods will need to be devised to prevent live progeny. Because of our adherence to standard culturing techniques, these methods are likely to be applicable in the context of the Lifespan Machine as well. Another is the use of temperature-sensitive sterile strains. However, these solutions may not always be appropriate, and in certain cases the reproductive span of individuals may persist into late ages, precluding the simple solution of transferring animals onto scanner only after they have ceased reproducing. The Lifespan Machine would not be very useful in such instances.

In the main text we demonstrate the use of the Lifespan Machine as an assay for survival under exposure to the oxidant t-BuOOH (Fig. 5 of main). Quite generally, our method can be employed to assess the impact of chemicals on lifespan. The inaccessibility of plates, however, requires that chemicals must persist at defined initial concentrations throughout an experiment, as they cannot be replenished.

- *Traits that may affect the acquisition and interpretation of lifespan data*
  There may be mutants whose behaviors makes automated analysis impractical; for example mutants that show no movement late in life, or mutants that crawl off the plates at a high frequency, or mutants that bag cryptically late in life. The possibility of these cases underscores the importance of quality control with the Worm Browser to detect them.

Acknowledgments

We would like to thank Joy Alcedo for providing the *hsf-1* and *glp-1* mutant strains, Becky Ward and Debora Marks for critical reading of our manuscript and Catalina Romero, Debora Marks and members of the Fontana lab for helpful discussions and encouragement throughout this

project. We would like to thank Tom Kolokotrones, Eric Smith, and Lee Jen Wei for discussions and statistical advice, and Mason Miranda our departmental IT specialist, for patiently meeting our unbounded needs for data storage. This work was funded by NIH grants R03 AG032481, R03 AG032481-S1, and R01 AG034994-01.